LINGUA

POSNANIENSIS

REVIEW OF GENERAL
AND COMPARATIVE LINGUISTICS

LXVI (2)

LINGUA
POSNANIENSIS

CZASOPISMO POŚWIĘCONE JĘZYKOZNAWSTWU
PORÓWNAWCZEMU I OGÓLNEMU

LXVI (2)

TABLE OF CONTENTS

Casting a new 'eye' on an elusive suffix: The use and origin(s) of Japanese dialectal suffix '*-me*'[1]

Étienne Baudel

École des hautes études en sciences sociales, Paris
etienne.baudel@gmail.com | ORCID: 0000-0002-6901-2742

Abstract: In several regions of Japan, various dialects exhibit a peculiar suffix *-me*, which mostly attaches to animal names, and seems to carry out various functions. The exact nature of its use is, to some extent, still unclear, since this *-me* is alternatively described as expressing endearment, familiarity or smallness, or, on the contrary, depreciation, fear or distance (which lead to some authors regarding it as two different homophonous suffixes). On a more grammatical scale, its relationships with definiteness, with plural and collective number and with animacy have also been questioned. Thus, this article aims at examining the different uses of *-me* in Japanese dialects, and its possible connection with other suffixes, in order to tackle the question of its origin(s).

Keywords: Japanese, dialectology, lexicology, suffixation, lexical semantics

Introduction

Although they are separated by hundreds of kilometres, several Japanese varieties or topolects reportedly exhibit a peculiar *-me* suffix that appears to be unused in the standard dialect of Tōkyō. This suffix usually attaches to animal names, as in: *ushi* 'cow' → *ushi-me*. However, its exact function(s) and origin remain somehow problematic, as well as its relationship with other homophonous suffixes.

An animal *-me* suffix has been observed in Japan since premodern times. For instance, it is mentioned as early as 1775 in Koshigaya's *Butsurui shōko* that people from the Sashima district of Shimōsa province (now Ibaraki prefecture) 'call [animals] by adding a *-me* mora to the end [of their names]', as in *mā-me* 'horse', *ka-me* 'mosquito' or *tonbo-me* 'dragonfly' (Koshigaya 1775: 1). A similar observation was quickly made about

---

[1] Many thanks to Igarashi Yōsuke, Matsukura Kōhei, Irène Tamba and Maxime Bonnet for their comments.

eastern Tochigi dialects by Ōta (1829: 472), then about Hachijō (Kondō 1855: 328). Later on, this suffix was gradually observed in other prefectures: Shiga in 1907, Ōsaka (Higashinari-gun) in 1922, Fukushima in 1932[2], Nara in 1936 (Tachibana 1936: 153), some parts of Ishikawa, Fukui, Saitama, Tōkyō and Kanagawa prefectures (Morishita 1981: 175), and, finally, in some parts of the Gifu prefecture (Yamada 1984: 1).

Overall[3], the reported distribution of the animal suffix -*me* in the Japanese archipelago can be mapped as represented on Figure 1[4]:
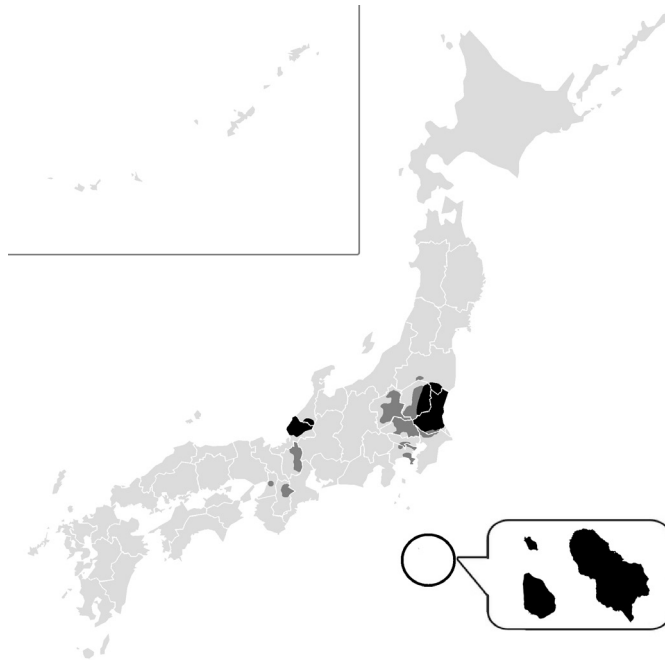
Figure 1: Map of the distribution of the -*me* suffix, with a zoom on the South Izu islands.
(The regions where an animal suffix -*me* was reported vivid in the 20th century are displayed in black, regions where it was reported moribund are displayed in dark grey.)

---

[2] According to the entry め of the *Nihon kokugo daijiten* (Shōgakukan 2001: 1069).

[3] According to Yamada (1984: 6), it is also not impossible to consider that traces of a comparable suffix are also attested in other places, since animal names ending in -*me* (and whose cognate do not end in -*me* in in standard Japanese) are attested on a few dialect maps; namely: Shimane (Izumo) *gyā-me* 'frog' (SJ *kaeru*) or *mokuro-me* 'mole' (SJ *mogura*), North Okayama *mokura-me* 'mole' (SJ *mogura*), and Miyazaki *kanchorome* and Kumamoto *ogami-me* 'praying mantis' (SJ *kamakiri*).

Similarly, a couple of animal names ending in -me can be observed in the *Gendai nihongo hōgen daijiten* (Hirayama et al. 1992-1994), such as Miyazaki *chocho-me* 'butterfly' (SJ *chō*), or Shimane *mokuro-me* 'mole'; and there are several similar occurrences in the *Nihon hōgen daijiten* (1989), such as Yamagata (Akumi-gun) cha-me 'cat' (SJ *neko*), Yamagata (Hidaka-gun) *ari-me* 'ant', Shimane (Yatsuka-gun) *kuchinā-me* 'snake' (SJ *kuchinawa*), etc. A list of such animal names can also be found in Hino (1946: 146-149).

However, as such, this data is probably too shallow an evidence to include those topolects to the scope of this article.

[4] This map was made using a Wikimedia blank background, and is mostly based on Morishita (1984: 175), with additional data from the other quoted sources.

In a nutshell, that distribution can roughly be divided into 3 main non contiguous regions, namely, from West to East and North to South:
1. parts of Fukui, Shiga and Ishikawa prefectures (with traces in neighbouring Nara and Ōsaka prefectures),
2. parts of Tochigi, Ibaraki and Fukushima prefectures (with traces in neighbouring Kantō prefectures),
3. the South Izu islands (Hachijō, Kojima, Aogashima).

In all those areas, this suffix seems to have been widespread until recent times. However, due to the pervasion of standard Japanese, it is now gradually falling out of use[5]. For instance, among the roughly 120 animal names in the six-volume *Gendai nihongo hōgen daijiten* (Hirayama et al. 1992-1994), only four varieties appear to have more than one occurrence of a *-me* suffix, namely: Hachijō (55 occurrences), Ibaraki (10), Shiga (8) and Fukui (6).

Nevertheless, its large distribution in non contiguous areas raises some questions:
1. Are all observations of dialectal *-me* occurrences the same phenomenon? If so, what function(s) does this suffix seem to have in the said topolects?
2. If this suffix is indeed a shared feature from those Japanese varieties, is it a shared archaism or a parallel innovation? More generally, what origin can be proposed for this suffix?

These two questions are to be the main considerations of this article, which will mostly focus on data from the South Izu islands (Hachijō, Kojima and Aogashima). However, since the details of this question exceeds by far the scope of this article, no definitive answer to the origin of *-me* will be provided in the second part, but rather preliminary hints that may serve as a starting point for future research.

1. The uses and functions of *-me* in Hachijō

1.1. The animal suffix *-me*

According to the specialist Morishita Kiichi (1979a; 1979b; 1981; 1984), there is no doubt that all of the aforementioned Japanese varieties exhibit one and the same suffix *-me*, rather than coincidental homophones. His definition of this suffix can roughly be summed up into 6 points:
1. *-me* is only used with animal names (1979a: 811),
2. it can be attached to all kinds of animal names, regardless of the animal's type (1979a: 790) or size (1979a: 813). However, different phonetic rules might restrict its use, depending on the variety (1979a: 812-813; 1979b: 141),

---

[5] See Morishita (1979a: 816) and Hayano (1991: 94) on its decline in Ibaraki and Tochigi prefectures. See Morishita (1984) on its decline in the Fukui prefecture, and (1979b: 139-142) on its decline on Hachijō island.

3. *-me* is more commonly used with familiar animals (1979a: 813), and possibly more common in appellative occurrences (1979a: 816; 1979b: 135),

4. there is an 'emotional difference' in the use or absence of *-me* (1979b: 135). Namely, it would express some kind of endearment (1979a: 789). Therefore, Morishita calls it 'the *-me* of endearment' (親愛の「メ」, *shin'ai no 'me'*),

5. *-me* could possibly express finer nuances as well, such as plural (1979a: 816) or diminutive (1981: 172-173), but those connotations are much less clear,

6. its function is essentially different from the homophonous depreciative suffix *-me* (which Morishita calls 罵詈の「メ」, *bari no 'me'* – lit. 'the *me* of invective') encountered in Middle Japanese, which also exists in most of those Japanese varieties, but usually attaches to human names and pronouns (1979a: 789; 1979b: 144).

This definition seems quite widely accepted. For instance, the idea that it is more common for familiar animal names and would express some kind of endearment was seconded by Yamada (1984: 3), Nakamoto (1980: 100), Iitoyo et al. (1984: 74) or, more recently, by Iannucci (2019: 94). However, it is probably not useless to check those assertions with the Hachijō data I compiled while preparing my PhD dissertation[6].

The first thing that can be noted is that the suffix *-me* is (or was) undoubtedly very productive in Hachijō, since it can be found in all kind of utterances, from all topolects, by all genders and all social classes including in the oldest sources[7]. Overall, it can be found in at least one quarter of the roughly 800 animal names I gathered so far. Namely, it can attach to words from all *strata* of Japanese lexicon alike, for instance: *hebi-me*[8] 'snake' (native), *chōchō-me* 'butterfly' (Sino-Japonic), *kokko-me* 'robin' (onomatopoeic), *horusutan-me* 'Holstein cow' (foreign loan). It also occurs both in the oldest inherited lexicon, such as *nuse-me* 'black kite' (OJ *nose*), as well as in recent loans such as *kingyo-me* 'goldfish' (a recent loan from Tokyo Japanese). However, *-me* appears to be more frequent in distinctly local words, for instance *hototogisu* 'lesser cuckoo' rarely occurs with *-me*, while its more local synonym *kokkenkokejō-me* does in most occurrences[9].

In addition, *-me* can attach on words of any length in Hachijō, for instance *ba-me* 'cow' (1 mora), vs. *dēgodanekurē-me* 'oriental greenfinch' (8 morae, lit. '*daikon*-seed eater'). However, it can be observed that it is usually avoided in compounds: for instance, *tori* 'bird' and *mushi* 'insect' systematically occur with *-me* when used as a bare root; but without suffix when they are second element of a compound, as in *katō-dori* 'streaked

---

[6] See the appendix of my PhD dissertation (Baudel: forthcoming) for the complete data.

[7] At this stage, the oldest occurrence I could find was *zoku-me* 'bull' in Takahashi (1802). A few years later, the famous Hachijō wordlist found in Ōta (ca. 1811), contains no less than 9 occurrences of *-me*. See Baudel: 2024 for more information on premodern Hachijō sources.

[8] For simplicity purposes, I use modified Hepburn transcription for both Hachijō and standard Japanese within this article (although the exact realisation of the written forms might locally differ), and Kunrei transcription for Middle and Old Japanese. For other varieties, I simply follow the quoted reference's conventions.

[9] Similarly, *dēgodanekurē-me* 'oriental greenfich' also has a more 'standard' variant *daikondanekurai* (found in Asanuma 1999), which has no *-me* and an uncontracted pronunciation of the vowel sequence /ai/.

shearwater', or *bun-mushi* 'beetle'. This rule bears almost no exception[10]; thus, an occurrence of *-me* after a compound might indicate that it is not synchronically analysed as such any more, for instance *shōto-me* 'turtle dove' (from **shiro+hato* 'white'+'dove'), *nyattori-me* 'chicken' (from *niwa+tori* 'garden'+'bird') or *mimizuku-me* 'horned owl' (from *mimi+tsuku* 'ear'+'owl').

In addition, one can observe that in Hachijō, *-me* can also attach to words that already end with the mora *me* (as in *same-me* 'shark' or *kame-me* 'turtle'); which, according to Morishita (1979a: 812) is not the case in some other varieties. However, the sequence *-me-me* seems to be avoided in longer words: for instance, **tsubame-me* 'swallow' does not occur (while both *tsubame* and *tsubakura-me* do), and *suzume-me* 'sparrow' is a rather rare form (*suzume* being much more common).

More generally, it can be noted that the occurrence of *-me* in Hachijō appears to be largely lexicalized, since many words (in fact, a majority) occur either only in their suffixed version, or only without suffix[11]. Besides, it can also be noted that *-me* cannot co-occur with an honorific suffix, which is most most visible in *kona-sama* 'silkworm' (in which *-sama* is now obligatory) and *nezumi-dono* 'mouse' (which also has a non-honorific variant *nezumi-me*). It can also be noted that *-me* often triggers syncope in suffixed words, causing a variety of sound change (for instance in **shirami* 'louse' > *shan-me*, *numi* 'flea' > *nun-me*, or *heijigara-me* > *heijigo-me / heiju-me* 'Japanese tit'), which might play a role in its lexicalization.

In this perspective, it can also be noted than in some rare but notable occurrences, *-me* seems to occur in free variation with an element of composition, such as:
- *tori* 'bird' (for instance in *hiyo-me ~ hiyo-dori* 'chick', *uno-me ~ uno-tori* 'cormorant', etc.),
- *yo* 'fish' (e.g. *tobi-me ~ tobi-yo* 'flying fish', *kingyo-me ~ kingyo-yo* 'goldfish'),
- *mushi* 'insect' (e.g. *uji-me ~ uji-mushi* 'maggot', *aka-me ~ aka-mushi* 'harvest mite'),
- another word (e.g. *goma-me ~ goma-fugu* 'tiger puffer'*, chōsen-me ~ chōsen-ushi* 'red cow' [lit. 'Korean cow']).

In all those occurrences the *-me* form seems to be more common, and it seems most likely that those forms result from a clipping (**u-no-tori+me → unome*; **chōsen-ushi+me → chōsen-me*); clipping being a very frequent phenomenon in Hachijō.

Due to these clippings, it can also be noted that in Hachijō, in synchrony, adding a suffix *-me* to a human name (such as a kinship term) or an adjective might give it an animal meaning, and more specifically make it refer to cows or to chickens. For instance, *ba-me* 'cow' can synchronically be analysed as *ba* '(grand)mother / aunt' + *-me* (probably from a clipping of **ba-ushi+me* 'mother-cow'). Similarly, there are forms such as

---

[10] Similar examples include *kubona-me* 'spider', vs *ye no kubona* 'house spider', *yama-kubona* 'wild spider'; *chaccha-me* 'warbler', vs *aka-chaccha* 'robin'; *ushi-me* 'cow', vs *chichi-ushi* 'milking cow', etc.

[11] So far, I found only four examples, namely: *ushi(-me)* 'cow', *zokku(-me)* 'bull', *nezumi(-me)* 'mouse', and *yagi(-me)* 'goat'. It is possible that this scarcity is due to a lack of data.

onnago-me 'young cow / hen' (lit. 'woman'+*-me*[12]; from *onnago-ushi+-me* and *onna-go-dori+-me*), *onokogo-me* 'young bull / rooster' (lit. 'man'+*-me*, from *onokogo-ushi+-me* and *onokogo-dori+-me*), *ko-me* '(cow) calf' (lit. 'child'+*-me* from *ko-ushi+-me*). This phenomenon was especially productive in the case of cows, as *-me* can attach to:

- a place name: *horusutan-me* 'Holstein cow', *chōsen-me* 'red cow' (lit. 'Korean-*me*'),
- an adjectival root: *kuro-me* 'black cow', *hēta-me* 'tired cow',
- a verbal noun: *tsukē-me* 'working cow', *hijike-me* 'skinny cow',
- any noun which acts like a qualifier: *konnyaku-me* 'spotted cow' (lit. 'konjac'+*-me*), *madara-me* 'pretty cow' (lit. 'elegant clothes' + *-me*).

In most of those occurrences, *-me* is in free variation with either *ushi* 'cow' or *zoku* 'bull', which seems to indicate that they all result from a productive form of clipping.

Finally, regarding the type of animals, it seems that there is no strong restriction in Hachijō. As a matter of fact, *-me* can attach to all kinds of animal names:

- wild and tamed (e.g. *kamo-me* 'wild duck', vs *ahiru-me* 'domestic duck'),
- positively and negatively perceived (e.g. *nekko-me* 'cat', vs *uji-me* 'maggot'),
- big and small (*kujira-me* 'whale', vs *ari-me* 'ant'),
- unique and in flocks (*pochi-me* 'doggy', vs *yuwashi-me* 'pilchards'), etc.

So far, the only tendency that I could observe is that while *-me* is almost ubiquitous in non-compound bird, reptile, mammal, insect[13] and crustacean names, it is more rare in fish names[14] and almost unused for shellfish[15].

Thus, in a nutshell, it can be said that Morishita's point 1 and 2 are quite strong in the case of Hachijō. However, the points 3-6 to appear much less clear. As a matter of fact, *-me* does not seem particularly more frequent in 'familiar' animal names than in rarer ones[16], and while both Nitta (2006: 131-133) and Morishita (1979a: 816) consider that the suffix *-me* might also be linked to the expression of plural or collective, it appears that this connotation does not show up in the data either, since I could find no statistical correlation between plural meaning and the use of *-me* (the singular use being slightly more frequent). However, it is possible that what Nitta and Morishita implied when mentioning a 'collective' or a 'familiar' meaning of *-me* is that this suffix could be more common for *generic* animal names rather than *specific* animal names (for instance in *fugu-me* 'pufferfish [in general]', vs *goma-fugu* 'tiger puffer', *jiri-fugu* 'brown-lined puffer', *bunpei-fugu* 'purple puffer', etc.). This last statement is true, but it actually does not say anything about the semantic value of the *-me* suffix *per se*, since (as illustrated with the example of pufferfish names) generic words are most commonly single-root words, while specific animal names are very often compounds; and, as showed earlier, *-me* tends to be avoided in compounds.

---

[12] It can also be noted that *onna-me* (lit. 'woman'+*-me*) is a common dialect word for 'cow' in Kyūshū, Kansai and some parts of Kantō (see Tachibana 1933). While a striking parallel, this form is however more probably an alteration of the more common *uname* (probably from *ushi-na-me* 'cow-GEN-female').

[13] I use 'insect' in a broad sense here; that is, including arachnids, worms, myriapods etc.

[14] For instance, some very common generic fish names such as *katō ~ katsū* 'bonito', *tē ~ chā* 'sea bream' or *shake* 'salmon' never take *-me*. Besides, the generic word for 'fish' *yo* does not either.

[15] The only clear exception seems to be *uni-me* 'urchin'.

[16] Cf, for instance, very common *hototogisu* 'lesser cuckoo', vs very uncommon *hya:daka-me* 'goshawk'.

Furthermore, Morishita's idea that *-me* would be more common in appellatives and express an 'endearment' meaning is not that clear either. As a matter fact, while *-me* does occur in some appellatives (as in *kan-me, kan'yo-me* 'kitty', or *pochi-me* 'doggy'), this use does not appear statistically more common than any other use in Hachijō, and is, in fact, overall quite rare. Similarly, while the interpretation of *-me* as a diminutive suffix (as suggested in point 5) could make sense on a diachronic perspective (*cf infra*); on a synchronic level however, there does not seem to be an opposition of size between suffixed and unsuffixed animal names, the two forms being used seemingly interchange-ably:

(1)   nezumi-**ME**=ga   anā   ake-tā=de
        mouse-**ME**=SBJ   hole.ACC   open-PST.RT=CIRC
    Since the mice made a hole... (Asanuma 1999: 78)

(2)   nezumi=ga   kabe=ni   anā   ake-tā=nte
        mouse=SBJ   wall=LOC   hole.ACC   open-PST.RT=CIRC
    Since the mice made a hole in the wall... (Asanuma 1999: 179)

Finally, while it is true that many occurrences of *-me* concern animals that are likely to receive some kind of endearment, such as cattle and pets, one could wonder if there truly is an endearment meaning in words such as *mushi-me* 'insect, pest', *uji-me* 'maggot' or *gejigeji-me* 'millipede', which are highly negatively-perceived animals, the latter even being used as an insult in Hachijō (cf. Asanuma 1999: 86). Besides some occurrences of *-me* clearly show a fearful or negative attitude toward the animal, perhaps showing some overlapping with the homophonous human depreciative *-me* (*cf infra*):

(3)   same-**ME**=ni   kam-are-te   tenpō d-ara
        shark-**ME**=A   eat-PASS-CONJ   armless be-PST.SS
    He is armless because he was bitten by a shark. (Asanuma 1999: 154)

(4)   nekko-**ME**=ga [...]   katsū-busho hikizuridashi-te bun-nige-tara
        cat-**ME**=SBJ   bonito.flake.ACC drag-CONJ   INT-run-PST.FIN
    The cat stole my bonito flake and ran away.

kondo   tsukamē-te buk-koroshi-te   yar-u
Next.time   catch-CONJ INT-kill-CONJ   do-PST.SS
    Next time I catch him, I'll kill him! (Asanuma 1999: 249)

Finally, it is interesting to observe that the suffix *-me* can occur even when talking about the inanimate representation of an animal, for instance in:

(5)   kono tsuru-**MĒ**   tsukur-o=wa   ezu-kya
        this crane-**ME**.ACC   prepare-RT=TOP   difficult-SS
    Making this [paper] crane is difficult. (Asanuma 1999: 47)

which means that synchronically, the use of *-me* is not necessarily linked with animacy in the speakers' mind.

In a nutshell, the animal *-me* suffix in Hachijō can be defined as follows:

1. *-me* is found after most animal names, regardless of the type of utterance,
2. it is used for all kinds of animals, with the notable exception of some fish and of all kinds of shellfish,
3. it is usually not found after compounds, and triggers frequent clippings
4. it cannot co-occur with another suffix, and tends to be avoided after the mora *me* in longer words (> 3μ),
5. its distribution is largely lexicalized, and appears to have undergone (at least partial) semantic bleaching.

Because of this semantic bleaching, the animal suffix *-me* appears to occasionally interfere to some extent with the homophonous human suffix *-me* (Morishita 1984: 183). Some authors like Naitō (1979: 108), Yamada (1984: 3), and Nitta (2006: 127-129) even consider that both could, to some extent, be considered two uses of the same phoneme.

Thus, it is probably not useless to take a closer look at the human uses of *-me*, in order to assess on a synchronic level to what extent the two appear to be linked or different; and on a diachronic level whether they share the same origin.

1.2. Other uses of *-me*

A human suffix *-me* also seems to be attested in many Japanese varieties, including the ones which also exhibit an animal suffix *-me*. This suffix is more difficult to observe than its animal counterpart and was possibly less productive. Plausible examples I could find[17] so far include: Shiga (Takashima-gun), *nana-me* 'child', Nara (Uchi-gun) *kogo-me* 'child', Fukishima *yaya-me* 'baby', Yamagata *gakki-me* 'brat', Ibaraki *oto-me* 'child', Tochigi *ani-me* 'big brother' (Nitta 2006: 143), Hachijō *appa-me* 'baby'. As can be seen in these occurrences, this suffix appears particularly common in kinship terms and has a very clear diminutive meaning, since it is mostly found after words that are used to denote or call infants.

In Hachijō, this human diminutive *-me* suffix does seem to have been quite productive, since it attaches to various lexical *strata*: *kodomo-me* 'child' (native), *chōnan-me* 'eldest son' (Sino-Japonic). It is used after several synonyms for 'children' (such as *kodomo-me* 'child', *appa-me* 'baby', *mago-me* 'grandchild'), especially after 'kin numerative'[18] appellatives, such as: (for boys) *tarō-me* '1st son', *jō-me* '2nd son', *sabō-me* '3rd son', *shō-me* '4th son', *gorō-me* '5th son' or (for girls) *nyoko-me* '1st daughter', *naka-me* '2nd daugther', *tego-me* '3rd daughter'; as well as their non-appellative equivalents: *chōnan-me* '1st son', *jinan-me* '2nd son, *sannan-me* '3rd son'; *chōjo-me* '1st daughter', *jijo-me* '2nd daugther', *sanjo-me* '3rd daughter', *yonjo-me* '4th son'. Interestingly, these series illustrate that *-me* is only found after the most common of these words, since the other forms of these series (such as *kusu* '4th daughter', *jīrō* '5th daughter', *kūrū* '6th daughter') are usually

---

[17] Unless stated, the following forms are quoted from the *Nihon hōgen daijiten* (Shōgaku tosho 1989).

[18] *Cf* Baudel: forthcoming a, for an overview of 'kin numeratives' in Hachijō.

not suffixed. Similarly, it can be noted that *shiko* 'great-grandchild' and *yashago* 'great-great-grandchild' are usually not suffixed either; which could either be due to their rareness, or to their perception as a compounds (that is, if this human *-me* has the same restriction as its animal counterpart).

Besides, it can be noted that in Hachijō, the human *-me* appears as less lexicalized than its animal equivalent, since it is appears completely facultative in most cases:

(6) kodomo-**ME**-ra=ga   heya=no nakā   hoger-o=wa
     child-**ME**-PL=SBJ   room=GEN=middle.ACC   scatter-RT=EMPH
   The children are scattering inside the room. (Asanuma 1999: 209)

(7) kodomo-ra=ga   urusa-ku-te   tsun-noboser-o=wa
     child-PL=SBJ   noisy-ADV-CONJ   INT-get.dizzy=RT=EMPH
   The children's noise is making my head ache. (Asanuma 1999: 149)

(8) mago-**ME**=wa   kawai-ke=nte   dēji=ni   sh-o=wa
     grandchild-**ME**-TOP   cute-RT=CAUS important=CIRC do-RT=SS
   My grandchildren are cute, so I cherish them. (Asanuma 1999: 209)

(9) mago=ga   bō-ku   nar-o=wa
     grandchild=SBJ   big-ADV   become-RT=EMPH
   My grandchildren are growing up. (Asanuma 1999: 209)

As shown from the example above, *-me* can co-occur with other suffixes, especially with the plural marker *-ra*, and the honorific marker *-do*, for instance in *asei-do-me* 'big brother'. In this context, it should be noted that the order of those suffixes appears to be fixed, since **-me-do* or **-ra-me* do not occur. I also found no occurrence of suffix-induced syncope with this suffix (on contrary to the animal suffix in **shirami > shan-me*, or *numi > nun-me*).

Beside kinship, another very common use of the human *-me* suffix in Hachijō and other varieties is in derogatory words and insults, such as *dongo-me* 'fool', *yakko-me* 'bastard', *fungī-me* 'idiot', etc. In a few Hachijō occurrences, such as *teren-me* 'liar', *yakkē-me* 'troublemaker', this *-me* occurs in free variation with *-mono* 'guy', 'person' and seems to work as an agentive meaning (*teren* meaning 'lie' and *yakkē* 'trouble'). It is likely that similarly to the animal *-me*, this use originates from a clipping (**teren-mono-me > teren-me*, *yakkai-mono-me > *yakkē-me*).

This derogatory *-me* is equivalent to a well-known derogatory or humbleness suffix *-me* (sometimes written as <奴>) which was very common in Late Middle and Early Modern Japanese, and in now given as old-fashioned in dictionaries and grammar books. This suffix is well attested at least since the *Heike monogatari* (early 14th century), in which it can attach to human and animal nouns and proper names alike:

(10) この程親しき奴めに   盗まれて候ふ
      kono hodo sitasiki yatu-**ME**=ni   nusum-are-te saurafu
      PROX INT close-RT guy-**ME**=AGT   steal-PASS-CONJ be-SS
   That so very dear fellow stole from us. (Imaizumi 1900: 505)

(11)   その仲綱めに   鞍   置け
       so=no Nakatuna-**ME**=ni   kura   ok-e
       MED=GEN Nakatsuna-**ME**=LOC   saddle   put-IMP
       Put a saddle on that Nakatsuna. (Imaizumi 1900: 497)

Occasionally, *-me* is interpreted as having an endearment meaning, especially when following a word referring to a child, for instance in the *Soga Monogatari* (14th century):

(12)   あの童めを   おとうと   おぼしめせ
       a=no wappa-**ME**=wo   otouto=to   obosi-mes-e
       DIST=GEN child-**ME**=ACC   little.brother=SIM   think-HON=IMP
       Consider this child like a little brother. (Anayama 1940: 117)

Later on, there are also many occurrences of *-me* after pronouns. For instance, in the *Arte da Lingoa de Iapam*, Rodriguez gives a short description of *-me*, which he describes as 'a particle that attaches to nouns and depreciates a lot' (1604: 162 *recto*), as in: <*vonoreme*> *onore-me* 'you (pej.)'*, <*aitçume*> *aitsu-me* 'that guy'*, <*coitçume*> *koitsu-me* 'this guy'. He also provides a long example sentence from a lost folktale (the *Kurofune Monogatari*)[19]:

(13)   nauo   vchixumeramo
       na^wo   uti-syu-**ME**-ra=mo
       ADV   interior-folks-**ME**-PL=COM
    Besides, even the people inside,

       corenofutariga   amari nurusani   ano yatçubara
       kore=no hutari=ga   amari nurusa=ni   ano yatu-bara
       PROX=GEN 2.ppl=SBJ   such leniency=LOC   DIST guy-PL
    The two of them, being so lazy, those guys,

       votocome mo,   vonagomemo   xŏdainai
       ^wotoko-**ME**=mo   ^wonago=**ME**=mo   syoodai na-i
       boy-**ME**=COM   girl-**ME**=COM   shape is.not-SS
    Neither the boy or the girl are any good. (Rodrigues 1604: 162 *recto*)

As can be seen in those examples, *-me* seems to be linked to definiteness in Middle Japanese, since it very often occurs after a deictic (*kore no, kono…*) or after a proper noun, which both behave like determinatives (Anderson 2004: 470). Thus, it usually does not occur in an indefinite context. Besides, it can be noted that at this stage, I found no occurrence of *-me* after a completely inanimate noun.

Interestingly, the use of *-me* with a pronoun is not common in contemporary Hachijō, and the only occurrence I could find so far is the premodern *unu-me* 'you (pej.)', which

---

[19] Interestingly, Rodrigues also provides an example of *-me* following an animal name: *inu-me* 'dog'.

occurs several times in Kakusō (1858 [1984:75]). However, as stated before, *-me* is quite commonly used after names in Hachijō, especially kin numeratives (which are most commonly used as appellative names), and it also appears to be rather rare in indefinite context, as illustrated by the examples below:

(14) wa=ga e=ni=wa   kodomo-ra ga   shōbu   ar-o=wa
      I=GEN house=LOC=TOP   child-PL=SBJ   many   be-RT=EMPH
    There are *many* children in my house. (Asanuma 1999: 117)

(15) kodomo-**ME**-ra=ga   warusā   shi-tā=nte sāg-o=wa
      child-**ME**-PL=SBJ   mischief.ACC   do-PST.RT=circ=TOP scorn-RT=EMPH
    I'll scorn *the* children: they did mischief. (Asanuma 1999: 103)

Similarly, the use of derogatory *-me* is more frequent in either appellative/vocative use or in definite context altogether, than in other kinds of utterances:

(16) urya   dongo   d-ara
      DIST.TOP fool   be-ss
    That guy is *a* fool.   (Asanuma 1999: 194)

(17) dongo-me=ga anshī   sogan   d-ā   kotō   sho?
      fool-**ME**=SBJ why such   be-RT   thing.ACC do-RT
    [*You*] fool, why do you do such things? (Asanuma 1999: 29)

Later on, this Late Middle Japanese suffix also became especially common in Early Modern Japanese following the humble first-person pronouns such as 某 *soregashi*, 拙者 *sessha* or 私 *watakushi*, in order to reinforce (sometimes ironically) their humbleness meaning. This can be seen for instance in Jippensha's picaresque novel *Tōkaidōchū Hizakurige* (1822 [1935: 31]):

(18) 下拙の私めが   相方の   おやまさんは
      gesetsu=no watakushi-**ME**=ga aikata=no   oyama-san=wa
      humble=COP I-**ME**=GEN2   partner=COP girl-HON=TOP
    The darling companion who will accompany my humble self...[20]

Similarly, the *-me* suffix is said to be especially common when talking about one's family or children, for instance in the Japanese translation of *The Last Days of Pompeii*:

(19) どうぞわたくしの息子めを   まもりくださいませ
      dōzo watakushi=no musuko-**ME**=o   mamori kudasai-mas-e
      please I=GEN son-**ME**=ACC   protect bestow-HON-IMP
    Please, protect my son. (Shibata 1951: 122)

---

[20]   The sentence is interrupted in the original.

It is most likely that this use is the origin of the kinship *-me* in Hachijō. However, interestingly, the use of *-me* with the first person pronoun is not attested in Hachijō, and it is quite clear in several examples that the kinship use of *-me* is not limited to one's own family in that variety:

(20) omi-ra=ga e=no   mago-**ME**=wa   soccha=n ik-ara
      you-PL=GEN house=GEN   grand.son-**ME**=TOP   MED=LOC go-PS.RT
    Your grandson went this way. (Asanuma 1999: 130)

(21) omi=ga nyoko-**MĒ**   wa=ga tarō=n
      you=GEN2 1st.daughter-**ME**.ACC I=GEN2=DAT
      yome=ni ke-nnō=ka
      wife=SIM give-NEG.RT=INTER
    Would you give *your* eldest daughter to my eldest son as wife? (Asanuma 1999: 89)

Thus, while it is less lexicalized and bound than the animal *-me*, the human *-me* suffix also appears to be undergoing some semantic bleaching in Hachijō. Interestingly, it can also be noted that this human suffix *-me* was also productive after names of (usually negative) supernatural beings in Hachijō, since there are several forms, such as *kinchi-me* 'ghost'*, oni-me* 'demon', *tenji-me* 'spirit', which rarely occur without suffix[21].

Perhaps more surprisingly, it also occasionally occurs (in a minority of occurrences) with names of meteorological phenomena, such as *tatsu-me* 'tornado', *kumo-me* 'cloud', and *niji-me* 'rainbow', in which a derogatory meaning is not perceptible any more. This use might somehow be ancient, since there are occurrences such as *arashi-me* 'storm' in the *kyōgen Setsubun*, dated from the end of the 16th century (Sasano 1956: 406)[22].

Finally, there is one last peculiar use of *-me* in Hachijō which makes it function like a bound noun[23], not unlike standard Japanese 者 *mono*. Namely, *-me* cannot occur as an independent word, but it can be qualified by a demonstrative, a proposition or an adjective:

(22) oso s-o=**ME**=ga ar-i
      lie do-RT=**ME**=SBJ be-NOM
    Since there are *people* who lie… (Kaneda 1998: 27)

(23) so=no waka-ke=**ME**-ra=ga
      MED=GEN young-RT=**ME**-PL=SBJ
    these young *guys*… (Kindaichi & Shibata 1967: 231)

---

[21] Nitta (2006: 127) also relates the same kinds of occurrences in Ishikawa.

[22] Alternately, it is also possible that in those few occurrences, human and animal *-me* conflated together, since *tatsu* originally means 'dragon', and rainbows in the Chinese mythology are said to be two-headed dragons.

Besides, there are also a couple of occurrences of facultative *-me* for body parts, namely: *hesso-me* 'navel' and *chinchin-me* 'penis', which probably show a recent extension from humans to human body parts.

[23] Ōshima (1984: 18) sums up that it is 'used as a formal noun. It may also function as a quasi-substantive particle.' (形式名詞として使われたもの。 さらに準体助詞的働きをする場合がある。)

In the two examples above -*me* acts like a classifier with a plural meaning, but this bound -*me* can also express contrast between two individuals, not unlike English *one*:

(24)   so=no ane=no hō=ga   kyoudē d-ō=**ME**=i
          MED=GEN big.brother=GEN side=SBJ   younger be-RT=**ME**=LAT
       The elder brother [said] to the younger *one*… (Kaneda & Naitō 2002: 28)

Besides, it can also take an animal or even an inanimate meaning:

(25)   ware=mo mitsuke-tā,   nobot-t-ar-o=**ME**
          I=COM find-PST.RT   climb-CONJ-be-RT=**ME**
       I also found some climbing *ones* [goats]. (Ōshima 1986: 69)

(26)   ki=no futō-ke=**ME**   hoso-ke=**ME** ja=nai darou
          tree=COP large-RT=**ME**   thin-RT=**ME** CIRC=NEG CONJECT
       The big *one* of the trees, not the thin *one*. (Ōshima 1986: 23)

(27)   seiro=wa   adan d-ō=**ME**   tsuka-i yar-ō,   omē-ra
          tree=COP how be-RT=**ME**   use-NOM do-PST.RT you-PL
       What *kind* of a steamer did you guys use? (Ōshima 1986: 45)

Finally, this -*me* can also have an exclamative function, or simply serve as a nominalizer for the proposition, not unlike Japanese *koto* こと or *no* の:

(28)   ō,   okkana-ke=**ME**=no
          yes   scary-RT=**ME**=TAG
       Yes, how scary! (Tokyo Metropolitan Board of Education 1986: 22)

(29)   meibo=ga ar-u-**ME**=ga
          register=SBJ be-RT$_2$=**ME**=ADV
       The thing is, I have a register, but... (Aoyagi 1973: 87)

(30)   tor-u=no=wa   oku=ni   hē-te   ar-o=**ME**
          take-RT$_2$=NOM=TOP deep=LOC   go.in-CONJ   be-RT=**ME**
       To collect [abalones], [one has] to go deep. (Nakamoto 1984: 160)

To the best of my knowledge, this use appears to be specific to Hachijō, and it might very well be a recent development[24]. More specifically, given the fact that Hachijō is prone to clippings, I suspect that it originates from a truncation of originally derogative *yatsu-me or *mono+-me (which can refer to humans and objects alike), which then underwent grammaticalisation semantic bleaching.

---

[24] Ōshima (1984: 18) explains that it in the 1980s, this form was most common amongst younger speakers, which might be a hint of its recent formation.

In this perspective, it is not impossible that other dialects exhibit a parallel relexification of the suffix *-me*. For instance, the *Kojima chihō no hōgen-shū* (Togawa 1966: 142) reports a noun '*me*' in Okayama prefecture, which is translated as 奴 *yatsu*, 人 *hito*, 者 *mono* (lit. 'guy', 'person', '-man'), and as *koto* in such syntagms as *ii me / ee me* = いいこと (Togawa 1966: 10, 27). However, no example is provided, and more research is needed in order to asses whether this word shares the same properties with its Hachijō counterpart.

- – In a nutshell, it is clear that regardless of their origin, there are at least three different uses of *-me* in Hachijō:
- – *-me₁* (as in *ushi-me*), a generic animal suffix that appears to have a very weak semantic value and to be highly lexicalized
- – *-me₂* (as in *kodomo-me, dongo-me*), a more rare but less lexicalized human diminutive / derogatory suffix
- – *=me₃* (as in *okkanake=me*), a nominalising element that lacks any clear semantic value, and probably originated from *-me₂* through clipping

Now, the question remains whether the first two suffixes *-me* share the same origin, and what hypotheses can be formulated on that matter.

2. The possible origins of *-me*

The dominant idea among the various authors that examined the dialectal *-me* suffixes (such as Nakamoto 1980: 100; Nitta 2006: 134-135) is that, although those two elements differ in synchrony, they are most likely sharing the same etymology. While this does seem to be a parsimonious hypothesis, it is also possible that those two forms were originally unrelated but interacted and conflated over the centuries. Besides, there are many possible candidates for their origin, which need to be examined one by one[25].

2.1. An origin as a derogatory suffix?

The first root we need to consider as a potential candidate for grammaticalisation is *mé⁺* 'eye' < OJ *me₂* (< PJ *may) which is quite remarkable for its polysemy and diversity of use. Without being exhaustive, we can roughly list the following meanings:

1. 'eye', 'eyesight', 'vision', 'judgement' (眼, 目),
2. 'object of vision', 'target', 'goal' (目),
3. 'dot', 'point' (目),
   → 'dots of die' → 'possibility', 'experience', 'moment' (目),
   → 'graduation', 'unit', 'item', 'weight', 'coin', 'value' (目, 匁),
4. 'small hole' (目),
5. 'joint', 'meeting point', 'intersection', '(biology) genus' (目),

---

[25] It must be noted that since my main source of data, Hachijō, is an accentless variety, it was not possible at this stage to take the pitch-accent into consideration.

6. 'bud, sprout' (芽),

7. 'face', 'expression', 'appearance'[26], 'grain (of a surface)', 'pattern' (目).

Furthermore, this root is said to have been grammaticalised in several ways in MJ (Kawaguchi 1998), and to occur synchronically as various suffixes (Kō 2004: 230-237):

1. (already in OJ[27]) as a suffix indicating the place, the moment or the result of the process of a verb root after a verbal noun (from the meaning 'point'):
   – 切る *kiru* 'to cut' → 切り目 *kiri-me* 'end / cutting mark' (lit. 'cutting point'),
   – 縫う *nuu* 'to sew' → 縫い目 *nui-me* 'stitch' (lit. 'sewing point'),

(by extension) as a suffix indicating the process itself after a verbal noun:
   – 落ちる *ochiru* 'to fall' → 落ち目 *ochi-me* 'decline' (lit. 'falling point'),
   – 控える *hikaeru* 'to hold back' → 控え目 *hikae-me* 'restraint' (lit. 'restraining point').

2. (roughly since the 11th century[28]) as a degree or approximation suffix after an adjectival root, not unlike English *-ish* (from the meaning 'appearance'):
   – 多い *ō-i* 'numerous' → 多め *ō-me* 'quite numerous',
   – 綺麗 *kirei* 'neat' → 綺麗め *kirei-me* 'very neat'.

If the adjectival root is monomoraic, *-me* attaches to the adnominal form instead:
   – 濃-い *ko-i* 'strong' → 濃いめ *ko-i-me* 'somewhat strong, strongish'.

3. (since the mid 15th century at least[29]) as an ordinal marker after a numeral compound (from the meaning 'item'): 一年 *ichinen* 'one year' → 一年目 *ichinen-me* 'first year'.

Besides, as shown by its meanings 'hole' and 'point', this *'me'* is also possibly etymologically linked to, or conflating with ma 間 'gap', 'interval', 'place' (Kawaguchi 1998: 266-267). Similarly, it might have interacted with the OJ and MJ volitional suffix *-m-*, which was used to create many de-adjectival verbs that were gradually reshaped into vowel verbs in *-eru* (first 下二段, then 下一段) in MJ; yielding verbal nouns in *-me*:
   – 高 *taka* 'high' → 高める *taka-meru* 'to raise' → 高め *taka-me* 'raising',
   – 確か *tashika* 'certain' → 確かめる *tashika-meru* 'to ascertain' → 確かめ *tashka-me* 'confirmation'.

All of those uses of *-me* are attested in Hachijō, e.g.:

1. *mūka* 'six days' (SJ *muika*) → *mūka-me* 'the sixth day'
2. *mushiru* 'to pluck' → *mushiri-me* 'plucking mark' (lit. 'plucking point')
3. *katē* 'hard' → *kata-me* 'quite hard'
4. *nagē* 'long' → *naga-meru* 'to lengthen/lie down' → *nagame* 'lengthening/lying down'

However, the category (3.) is overall quite rare and possibly unproductive in Hachijō, as I found no compelling occurrence after a monomoraic adjective or after a Sino-Japanese adjective. Still, one can note the striking structural similarity in synchrony between

---

[26] In which *me* is often said to have conflated with 見え *mié* 'appearance', the *ren'yōkei* of the OJ verb 見ゆ *miyu* 'to be visible' (modern *mieru* 見える). See Martin (1975: 830-831).

[27] For instance, *MYS,* XII, 2967 already has 縫目 *nupi-me*, OJ form of 縫い目 *nui-me*.

[28] For instance in *The Tale of Genji*, 26 (*Tokonatsu*), 細目 *hoso-me* 'rather thin' (Murasaki 1927:259).

[29] References dictionaries such as the *Daigenkai* (Ōtsuki 1939: 568) and the *Daijiten* (Heibonsha 1935: 294) quote a first occurrence in a commentary of the *Shiki* dated from the Bunmei era (1469-1487).

the -me in standard Japanese koi-me 'strongish', and the -me in Hachijō okkanake=me 'how scary'; both occurring after the adnominal form of an adjective.

Among all those forms and meanings, the 'face', 'appearance'[30] one (7. in the list above) seems to be most plausible candidate for a grammaticalisation as a derogatory suffix, if one assumes that it was originally used in insults and subsequently reanalysed as a derogatory suffix[31], e.g.:

– aho 'fool' → aho-me *'idiot-face' → (reanalysed) aho + -me 'idiot-DER'

This derogatory meaning would then have been extended to express humility or self-contempt and would have been subsequently extended to talk with humility about one's children. Finally, the suffix would have been reanalysed as a diminutive, allowing its extension probably first to cattle animals, then to all animals in some varieties:

– 'face' → contempt → self-contempt → endearment → animality

This hypothesis seems plausible, but it supposes that the 'face' meaning was lost quite early on, since it is not perceptible any more already in the first occurrences, especially the ones that follow a proper name. However, one problem with this hypothesis is that whereas the endearment meaning is already attested very early on (as shown in the example (12) from the *Soga Monogatari*), the self-contempt one is not very common at that time, and is totally absent from some varieties like Hachijō.

2.2. An origin as a directive / focus marker?

Another possibility is to consider that the derogatory meaning of -me is secondary and that it was originally a directive or a focus marker. As a matter of fact, as stated earlier, the suffix -me is attested after a deictic in a large number of older occurrences, and is said to be especially productive in this use in some varieties such as the Sendai dialect from Miyagi (Doi 1919: 76).

Besides, several Japanese dialects[32] have an element -me that works as a directional marker, as in oki-me 'off-shore', higashi-me 'to the east', nishi-me 'to the West', and a homophonous adverb me seems to exist in at least some Ryukyuan languages[33]. Thus, it is possible to suppose that a *me deictic element (possibly originating from the meaning 'target', 'goal' of the root 'eye') was used in prehistoric Japanese with a meaning such as 'there', 'over there', and was reinterpreted as derogatory suffix because of it frequent use in apostrophe:

– *ano aho me 'that fool *over there'* → (reanalysed) ano aho-me 'that fool-DER'

While not completely implausible, this hypothesis remains quite speculative, since there is no proof that the directive element found in those varieties is ancient. Alternate-

---

[30] Not unlike like in English -face in fatface, prickface, jerkface; French tête de ..., etc.

[31] Theoretically, the meaning 'value' could also be a potential candidate, as in:
aho 'fool' → aho-me *'idiot-worth' → (reanalysed as) aho + -me 'idiot-DER'
However, this meaning is likely too recent and too limited to have played any role in this grammaticalisation.

[32] Namely, Yamaguchi prefecture (Shigemoto 1937 [1976]: 71; Yamanaka 1967: 166), Tsushima island (Takiyama & Yanagita 1944: 94) and Kimotsuki-gun, Kagoshima prefecture (Nomura & Yanagita 1942: 75).

[33] For instance in Taketomi island (Yaeyama). See (Maeara et al. 2011: 1206).

ly, it could very well be a rebracketing of the approximative suffix found after adjective roots, since the *-me* in *taka-me* 'high-ish' or *hiku-me* 'lowish', is very easily reanalysable as a kind of a locative ('on the high side' → 'facing up', 'on the low side' → 'facing down'), which could have led to its extension to direction names.

Quite in the same way, it is also possible to suppose that the original meaning of the derogatory *-me* was some kind of a focus particle, originally used for emphasis or contrast, not unlike English *very*. As a matter of fact, a *me* assertive particle (similar to standard よ) is found in some Ryukyuan languages, such as Yaeyama Ishigaki (Miyagi 2003: 1116), and a focus particle me_2 is said to occur in one occurrence of Eastern Old Japanese:

(31) 須流河乃祢良波   苦不志久米阿流可
       Suru^{n}ga-no$_2$ ne$_1$-ra pa   kupusi-ku=**me$_2$** ar-u=ka
       Suruga=GEN cliff-PL=TOP   be.longing-ADV=FOC be-RT=EMPH
   How I long / for the cliffs of Suruga... (MYS, XX, 4345, 4-5)

Besides, this hypothesis would work quite well with the oldest attestations of *-me*:
– *kono [...] yatsu-me* 'that *very* fellow'
– *sono Nakatsuna-me* 'that *very* Nakatsuna'
– *ano wappa-me* 'that *very* child'

Then, this putative *-me* could have specialised as a humble 1st person marker meaning 'me *personally*' in a devoted way, and, since the focalisation after a deictic was often negative, it would have gradually become a derogative marker, leading to its specialisation in animal names:
– focus → contempt / self-contempt → animal

While this hypothesis is also not completely implausible and would explain the high frequency of *-me* with deictics, it is also very shallow. As a matter of fact, EOJ *më* is a *hapax legomenon*, and might very well be a mere graphic variant of OJ *mo$_2$* due to phonetic peculiarities of the Suruga OJ variety (see Kupchik 2023: 31-32). Besides, the assertive particle *me* in Ishigaki does not have the same semantic and syntactic properties as a focus particle, and is most likely unrelated to the derogatory *-me*.

2.3. An origin as a composition element?

In addition to all those homophonous roots[34], it has also been noted for a long time (for instance by Suzuki 1816: 1-2) that an element *-me* was also very common common in bird names such as 雀 *suzume* 'sparrow', 鴎 *kamome* 'seagull' or 燕 *tsuba(kura)me* 'swallow', although it is usually not displayed in the spelling, all those birds having there own *kanji*. The same was also noted for fish names, in which this is even more common, for instance in 油女 *aburame* 'fat greenling', 泥目 *dorome* 'gluttonous goby' or 平目

---

[34] There is yet another homophonous root *me* (藻) meaning 'seaweed' (for instance in *waka-me* 'edible brown seaweed') which will not be discussed here since it is highly unlikely to have played a role in the emergence of the animal and human suffixes.

hirame 'flounder', in which the *-me* element is usually graphically visible and noted with a phonetic use of either <女> 'woman' or <目> 'eye'.

Besides, it was noted as early as 1775 (Koshigaya 1775: 1), and more recently by Maeda (1988: 27-29) that those animal names ending in *-me* used to be even more common in Old and Middle Japanese, with such bird names as 小雀め *kogarame* 'willow tit' (now 小雀 *kogara*), 鸊鷉 *uyome / iyome* 'grebe' (now 鸊鷉), 機織め *hataorime* 'weaverbird' (now 機織鳥 *hataori-dori*), 山雀め *yamagarame* 'varied tit' (now 山雀 *yamagara*); and fish names like 赤女 *akame*[35] 'sea bream' (now 鯛 *tai*) or 口女 *kuchime* 'young striped mullet' (now 鯔 *bora* or *nayoshi*). However, to the best of my knowledge, such an element *-me* does not seem attested in Ryūkyū bird and fish names.

This lead several scholars to suppose that there were yet two other roots **-me* in Old Japanese or pre-Old Japanese, respectively meaning 'bird' and 'fish'. Thus, several *koku-go* dictionaries, such as the *Daigenkai* (Ōtsuki 1939: 568) and the *Daijiten* (Heibonsha 1935: 294) came to treat those two *'me'* elements as separate entries.

Remarkably, several of those animal names seem to be analysable as semantically and structurally sound compounds, namely:

- *suzu-me* = 'chirping sound[36]' + *-me* = 'chirping bird'
- *tsuba-kura-me* = 'wing' + 'black' + *-me* = 'black-winged bird'
- *hataori-me* = 'weaving' + *-me* = 'weaver bird'
- *kogara-me* = 'small + tit[37]' + *-me* = 'small tit bird'
- *yamagara-me* = 'mountain + tit' + *-me* = 'wild tit bird'

And for fish:

- *abura-me* = 'oil, fat' + *-me* = 'fat fish'
- *doro-me* = 'mud' + *-me* = 'mud fish'
- *hira-me* = 'flat' + *-me* = 'flat fish'
- *aka-me* = 'red' + *-me* = 'red fish'

The words *kamome* (OJ *kamame*), *uyome* and *kuchime* are more difficult to analyse, but it is not impossible that the first two are linked with other aquatic bird names (namely *kamo* 'duck' and *u* 'cormorant'), while the second could be either a clipping of compound containing 口 *kuchi* 'mouth', or perhaps a compound with an unvoiced variant of the element *kuji-* found in 鯨 *kujira* 'whale' and 久慈目 *kujime* 'spotbelly greenling'.

Remarkably, while those words are all seemingly analysable as compounds, *-me* does not seem to have the same semantic meaning in all of them. Namely, in the case of *kogara-me* and *yamagara-me -me* is semantically superfluous, since tits are birds by definition. Thus, in these two words, *-me* seems to be a facultative suffix, which may be acting like a categoriser. On the contrary, *-me* appears to be a necessary component conveying the [animal] meaning in the other words, especially in all fish names.

---

[35] The word 赤目 *akame* still exists in contemporary Japanese, but means 'red mullet', which is a visually quite similar but much less common fish.

[36] According to Suzuki (1816: 1). This hypothesis seems to be corroborated by the various dialect forms *chichi(-me), chinchin(-me), choncho(-me)*, etc.

[37] This root is not found in isolation, but seen in several compounds: *shijū-gara* 'black tit', *eboshi-gara* 'tufted tit', *hi-gara* 'coal tit', *hige-gara* 'bearded tit', etc.

In addition, it can be noted that several of those words frequently occur without *-me* in some varieties, especially in compounds, for instance[38]:

– *tsubakura / tsubakuro* 'swallow' (common)
– *niwa-suzu* 'garden sparrow' (Chiba, Shimane)
– *shiro-gamo / shira-gamo* 'white gull' (Kumamoto, Shimane)
– *abura* 'fat greenling' (Kagawa)
– *de-bira* 'flounder' (Kagawa, Hiroshima, Ehime)

The fact that *-me* seems to be less common in compounds is a remarkable parallel with Hachijō. Besides, it is interesting to note that, if the analysis of *-me* as an element of composition is correct, the modern form *tsubame* 'swallow' is most likely the result of a clipping of *tsuba-kura-me*, since 'winged bird' is by no means a specific bird name.

Based on all these elements, it was theorised by several scholars (such as Tachibana 1933: 35-36; Kindaichi 1973: 172) that the *-me* element in birds and fish names was likely to be a word meaning 'animal' that was used as a suffix or an element of composition, and was gradually lost and lexicalized in standard Japanese, but was preserved in more conservative varieties as a facultative suffix.

In this case, the derogatory meaning would be secondary, and due to a pejorative comparison of the suffixed word with animals, namely:

– 'animal' → contempt → self-contempt

This hypothesis seems perhaps more plausible than the preceding ones. However, it is also quite speculative, given the fact that occurrences of the derogatory suffix *-me* precede the occurrences of the animal *-me*. Besides, one could still wonder about the origin of this putative animal marker **me*.

2.4. An origin as a collective marker?

In this regard, a hypothesis was already well established in the Edo period (Suzuki 1814: 2), claiming that the *-me* that is found in animal names was originally a collective suffix 群れ *-mure* 'flock'. In this case, *-me* could be cognate with the Ryukyuan plural suffix *-buri / -mmi* (Shimoji 2022: 506), with Taketomi *-me* (Maeara et al. 2011: 1206), and with a plural *-me* found in a few Kyūshū topolects such as Kuboizumi:

(32)   koton-**ME**=nya   kimon tsukut-te
       child-**PL**=DAT.TOP   clothes prepare-CONJ
    Preparing clothes for the children... (Kindaichi & Shibata 1966: 126)

More precisely, the meaning 'flock' would have been used for animals as a collective, and for humans as a derogatory plural (as in English 'bunch of'), leading to its split into two suffixes:

– 'flock'   → human plural → contempt → self-contempt → endearment → animal plural → animality

---

[38] All forms are quoted from the *Nihongo hōgen daijiten* (Shōgakukan 2003).

This hypothesis has several qualities. First of all, it is quite satisfactory in explaining with one form the origin of the two suffixes. Besides, it would make sense for gulls, sparrows and swallows (which most often move in flocks) to be referred to as a collective, which is the case in all occurrences of those words in the *Man'yōshū*. However, the shift **-mure > *-me* would require a rather irregular phonetic development quite early on in the history of the language, and would therefore need further research in order to be confirmed or infirmed.

2.5. An origin as a diminutive?

Finally, there is one last hypothesis which might be worth considering, namely that the original meaning of *-me* in Japanese varieties was a diminutive marker. As a matter of fact, as shown for instance by Jurafsky (1996), diminutives can come to express a variety of meanings beside smallness, among which endearment or familiarity, contempt, humility, specification and approximation, all meanings that, as we saw, can be expressed by the different *-me* suffixes in Japonic varieties. This can also be seen in Japonic for instance with the different meanings of Miyako diminutive *-gama*[39]:

(33)   ba=ga ffa-gama=a   kanasï-munu
          I=GEN child-DIM=TOP lovely-thing
     'My little child is lovely' (Michinori 2008: 247)

(34)   uma=nu   junsja-gama
          MED=GEN   policeman-DIM
     'That *bloody* policeman there...' (Michinori 2008: 247)

Besides, Antonov (2007: 195-198) showed that the diminutive *-ra* in Old Japanese was also linked with the expression of plural and locative, and Pellard (2010: 140) indicated that the diminutive in Ōgami was also frequently used with deictics in order to indicate 'imprecise direction', which could explain several of the specific uses that were discussed in the above.

In addition, as noted by Jurafsky (1996: 568), diminutives are also especially common in animal names, and can very easily get lexicalized. Thus, we can assume that the fossilised animal names in *-me* are lexical traces of a former diminutive suffix that was only preserved in some varieties, and which was specialised in standard Japanese as a derogatory suffix when following a noun, and (possibly) as an approximation suffix when following an adjective. A good argument in favour of this hypothesis could be the existence of a diminutive *-be* in some eastern varieties (already noted by Morishita 1981: 176-178), which could easily be a cognate of *-me*, as well as the existence of several dialect forms in which *-me* is replaced by the more common diminutive suffix *-ko*, as in[40]:

---

[39] As shown by Jarosz (2015), there is as least one lexicalized bird name in *-gama*: ssumi:-gama 'white-eye'.

[40] All forms are quoted from the *Nihongo hōgen daijiten* (Shōgakukan 2003).

- *tsuba-me* 'swallow' (standard)   = *tsunba-ko* (Chiba)
- *suzu-me* 'sparrow' (standard)   = *chichi-nko* (Ōita)
- *abura-me* 'fat greenling' (standard)   = *abura-ko* (Yamagata)
- *yama-me* 'masu trout' (standard)   = *yama-nko* (Hyōgo)
- *dō-me* 'round herring' (Hyōgo, Tottori, Shimane) = *don-ko* (Ishikawa)

Similarly, it can be noted that the Hachijō of Japanese *kamome*, *kabuna(-me)* exhibits the Eastern Old Japanese diminutive *-na*.

In this case, there is still to explain the origin of such a diminutive. In this perspective, a good candidate would probably be the homophonous root *mé*⁺ < Old Japanese *me₁* (< PJ *mya), meaning 'woman' (女), 'female' (雌), as already suggested by Nakamoto (1980: 100) and Martin (1975: 831). As a matter of fact, as shown again by Jurafsky (1996: 536), the concept of 'girl' is frequently associated with smallness or youth, especially in Japan[41]. For instance, it can be observed already in Old Japanese that *me* serves as a diminutive prefix, for instance in 女滝 *me-daki* 'small waterfall' (lit. 'girl waterfall'), vs 雄滝 *wo-daki* 'large waterfall' (lit. 'boy waterfall'), or 女波 *me-nami* 'small wave' (lit. 'girl wave'), vs 男波 *wo-nami* 'large wave' (lit. 'boy wave').

In compounds, *-me* also regularly occurs in parallel with *-ko* 'child' in Old Japanese:
- *oto-me* 'girl' vs *oto-ko* 'boy'
- *musu-me* 'daughter' vs *musu-ko* 'son'
- *iratsu-me* 'lass', vs. *iratsu-ko* 'lad'
- *hi-me* 'princess' vs *hi-ko* 'son'

Given the fact that *-ko* grammaticalised as a diminutive suffix in many Japanese varieties, it is not implausible to suppose that *-me* underwent the same grammaticalisation as *ko* (either by parallel development or as the result of a chain shift), but specialised in animal meanings or derogatory diminutive due to usually inferior position of women in the traditional Japanese society[42]. Besides, given the fact that *-ko* often marks an inanimate diminutive (as in *furi-ko* 'pendulum', *hashi-go* 'ladder'[43]), it is also not impossible that *-me* specialised early on as an *animate* diminutive, which would have made it specialisation in animal names even easier.

If this hypothesis is correct, *-me* would be comparable with the English suffix *-ster*, which was originally feminine in meaning, got fossilized in a few animal names (such as *lobster*), and went on to become a derogative agent suffix, for instance in *gangster, scamster, prankster* etc.:

- 'woman' → (animate) diminutive   → animality
                                          → endearment
                                          → contempt / self-contempt

[41] For example, the Yamagata hōgen kenkyū-kai (1970: 603) relates that the word *me* (etymologically 'female') now fully means 'child' in this variety.

[42] If one needs examples of the perception of women in ancient East Asia, one can for instance observe that the *kanji* radical <女> is also often pejorative, for instance in compounds such as <姸> 'quarrel' or <姦> 'wicked'.

[43] Compare the Chinese suffix <子> *-zi* for a similar phenomenon.

This hypothesis is, in my opinion, the most economic one, since it manages to bring back almost all meanings of *-me* to one single original meaning. However, it does remain quite shallow, given the fact that this putative diminutive does not occur in Old Japanese. Thus, at this stage, more research is needed in order to confirm or infirm this hypothesis.

Conclusion

In this article, we have showed that several Japanese dialects, and especially the varity of Hachijō exhibit a peculiar suffix *-me*, which raises a lot of questions, both synchronically and diachronically.

Synchronically, three main uses of *-me* can be distinguished in Hachijō:
1. a generic animal suffix, which looks quite lexicalized,
2. a less bound diminutive suffix, that is specialised in kinship terms and insults,
3. a nominalising element that lacks any clear semantic value,

However, it seems plausible that all of those elements share the same origin.

Finally, I have showed that there are several possible scenarios in order to explain the origin of these suffixes, namely:
1. 'face'　　　　　　　　→ contempt → self-contempt → diminutive → animality
2. *focus　　　　　　　　→ contempt / self-contempt → animality
3. *'animal'　　　　　　　→ contempt → self-contempt
4. 'flock'　　　　　　　　→ human plural → contempt → self-contempt → endearment
　　→ animal plural → animality
5. 'woman'　　　　　　　→ diminutive　 → animality
　　→ endearment
　　　　　　　　　　　　　　→ contempt / self-contempt

At this stage, more research is needed in order to disentangle these possibilities. Especially, more historical evidence from Old and Middle Japanese will have to be considered for further research, as well as more comparative evidence from other Japonic varieties.

References

Anayama Takamichi (ed.). 1940. *Soga Monogatari: Ōdōbon* [The tale of Soga: Book from the royal palace]. Vol. 2. Tōkyō: Iwanami shoten.

Anderson, John M. 2004. On the grammatical status of names. *Language* 80(3). 435-474.

Antonov, Anton. 2007. *Le rôle des suffixes nominaux en* r *dans l'expression du lieu et de la direction en japonais.* Paris: Inalco. (Doctoral dissertation.)

Aoyagi Ayako. 1973. Hachijō-jima no shizen kaiwa [Natural conversations from Hachijō island]. *Fīrudo no ayumi: Seikatsu-go kenkyū no kiroku* [Language in real life: Bulletin of the Research group on Language in areal life] 3. 103-114.

Asanuma Ryōji. 1999. *Hachijō-jima no hōgen jiten* [Dialect dictionary of Hachijō island]. Tōkyō: Asahi Shimbun Publishing Service.

Baudel, Étienne. 2024. Adopting a Philological Approach Toward *Chishi* (地誌). In Calvetti, Paolo (ed.), *Philological and linguistics analysis working together: Exploring language diversity in Japan*, 31-52. Venice: Ca'Foscari.

Baudel, Étienne. 2025. Dénommer les individus selon leur rang dans la fratrie: Les noms ordinaux dans les îles Izu du sud. In Couralet, Stéphane & Labrune, Laurence & Chieko Shirota (eds.), *Linguistique coréenne et japonaise: Dynamiques contrastives*, 253-265. Bordeaux: Presses universitaires de Bordeaux.

Baudel, Étienne. [forthcoming]. *Shima-kotoba: A synchronic and diachronic study of Hachijō*. Paris: EHESS. (Doctoral dissertation.)

Doi Yatsue. 1919. *Sendai hōgen-shū* [Collection of Sendai dialect]. Sendai: Doi Yastue.

Hayano Shingo. 1991. Tochigi Ibaraki ryō-hōgen no setsubi-go: 'me', 'ko' / 'bo / bō' ni tsuite [The suffixes *-me, -ko, -bo* and *-bō* in the dialects of Tochigi and Ibaraki]. *Nagoya hōgen kenkyūkai kaihō* [Report of the Nagoya Dialect Research Association] 8. 77-94.

Heibonsha (ed.). 1935. *Daijiten* [Large dictionary]. 264. Tōkyō: Heibonsha.

Hino Iwao. 1946. *Hyūga hōgen ronkō* [Essay on the Hyūga dialect]. Vol. 1. Miyazaki: Jōdai Hyūga kenkyūjo.

Hirayama Teruo, Ōshima Ichirō, Ōno Makio, Kuno Makoto, Kuno Mariko, Sugimura Takao (eds.). 1992-1994. *Gendai Nihongo hōgen daijiten* [Large dictionary of contemporary Japanese dialects]. Tōkyō: Meiji shoin.

Iannucci, David. 2019. *The Hachijō language of Japan: Phonology and historical development*. Mānoa: University of Hawai'i. (Doctoral dissertation.)

Iitoyo Kiichi. 1959. Hachijō-jima hōgen no gohō [Hachijō island dialect grammar]. *Kotoba no kenkyū: Kokuritsu Kokugo Kenkyūjo Ronshū* [Linguistic Research: Collection of Articles from the National Institute for Japanese Language and Linguistics] 1. 215-232.

Iitoyo Kiichi, Hino Takezumi, Satō Ryōichi (eds.). 1984 [1998]. *Kōza hōgengaku: Kantō chihō no hōgen* [Courses in dialectology: The dialects of the Kantō area]. Tōkyō: Kokusho kankōkai.

Imaizumi Teisuke (ed). 1900. *Heike Monogatari kōgi* [Lecture of the *Heike Monogatari*]. Vol. 2. Tōkyō: Seishidō.

Jarofsky, Daniel. 1996. Universal tendencies in the semantics of the diminutive. *Language* 73(3). 533-578.

Jarosz, Alexandra. 2015. *Nikolay Nevskiy's Miyakoan dictionary: Reconstruction from the manuscript and its ethnolinguistic analysis*. Poznań: Adam Mickiewicz University. (Doctoral dissertation.)

Jippunsha Ikku. [1822] (1935). *Tōkaidō chū Hizakurige* [Hizakurige in the Tōkaidō]. Ed. by Sankyō shoin henshū-bu. Vol. 2. Tōkyō: Sankyō shoin.

Kakusō Kizan [1848] (1984). *Yatake no nezamegusa* [The waking-up grass of Hachijō]. Ed. by Nakada Norio. 1984. Tōkyō: Benseisha Bunko.

Kaneda Akihiro 1998. *Aogashima-hōgen* [The Aogashima dialect]. Chiba: Chiba Daigaku Ryūgakusei Sentā.

Kaneda Akihiro & Naitō Shigeru. 2002. *Hachijō hōgen no ikita kotoba: Minwa, densetsu, danwa: taiyaku tsuki* [The surviving words of the Hachijō dialect: Folktales, Traditions, Conversations: With mirror translation]. Tōkyō: Kasama shoin.

Kawaguchi Junji. 1998. 'Me' no bunpōka wo megutte [The grammaticalisation(s) of *me*]. *Keiō gijuku daigaku geibun gakkai* [Keio University Arts and Letters Society]. 74. 253(106)-268(91). 253-268.

Kibe Nobuko (ed.). 2013. *Shōmetsu kiki hōgen no chōsa hozon no tame no sōgō-teki kenkyū: Hachijō hōgen chōsa hōkokusho* [General studies for research and conservation of endangered dialects in Japan: Research report on Hachijō dialects. Tachikawa: NINJAL.

Kindaichi Haruhiko. 1973. *Kotoba no saijiki* [Language calendar]. Tōkyō: Shinchōsa.

Kindaichi Haruhiko & Shibata Takeshi. 1966-1967. *Zenkoku hōgen shiryō* [Dialect materials from all over Japan]. Tōkyō: NHK.

Kō Kisei. 2004. *Gendai Nihongo no setsubiji kenkyū* [Research on the suffixes in contemporary Japanese]. Hiroshima: Keisuisha.

Kondō Tomizō. 1855 [ed. 1964]. *Hachijō Jikki* [The true account of Hachijō], vol. 1. Ed. Hachijō Jikki Kankōkai. Tōkyō: Ryokuchi-sha.

Koshigaya Gozan. 1775 [ed. 2015]. *Butsurui shōko* [Animal names], vol. 2. Ed. by NINJAL: (https://www2.ninjal.ac.jp/textdb_dataset/brsk/brsk-002.html) (Accessed 2025-01-30.)

Kupchik, John. 2023. *Azuma Old Japanese: A comparative grammar and reconstruction*. Boston/Berlin: De Gruyter/Mouton.

Maeara Tōru, Hateruma Eikichi, Takamine Hōyū, Irisato Teru (eds.). 2011. *Taketomi hōgen jiten* [Dictionary of Taketomi dialects]. Ishigaki: Nanzan sha.

Maeda Tomiyoshi. 1988. Bunken kokugo-shi to hōgen: Setsuji o chūshin to shite mita [Written history of the Japanese language and dialects: Focusing on affixes]. *Hōgen kenkyū nenpō* [Annual report on dialect studies], vol. 30, 17-31. Hiroshima: Hiroshima Dialect Research Institute.

Martin, Samuel. 1975 [2003]. *A reference grammar of Japanese.* New edition. Honolulu: University of Hawai'i Press.

Michinori Shimoji. 2008. *A grammar of Irabu, a Southern Ryukyuan language.* Canberra: The Australia National University. (Doctoral dissertation.)

Miyagi Shinyū (ed). 2003. *Ishigaki hōgen jiten* [The dialects of Ishigaki]. Naha: Okinawa taimusu sha.

Morishita Kiichi. 1979a. Dōbutsu-mei ni tsuku setsubigo 'me' ni tsuite [The suffix -me that attaches to animal names]. In Tanabe hakase koki kinen kokugaku ronshū henshū iinkai (ed.) *Tanabe-hakase koki kinen kokugo joshi jodōshi ronsō* [Collection of studies of Japanese particles and auxiliaries in honour of Pr Tanabe's 70th birthday], 789-817. Tōkyō: Ōfū-sha.

Morishita Kiichi. 1979b. Hachijō-jima ni okeru setsubigo 'me' no kinō ni tsuite [The function of the suffix -me in the dialects of Hachijō island]. *Iwate idai kyōyō-bu kenkyū nenpō* [Iwate Medical University, College of Liberal Arts Research Annual Report] 14. 134-144.

Morishita Kiichi. 1981. Shin'ai hyōgen no setsubigo ni tsuite 'me' o chūshin ni shite [The suffixes expressing affection: Focussing on -me]. *Iwate idai kyōyō-bu kenkyū nenpō* [Iwate Medical University, College of Liberal Arts Research Annual Report] 16. 172-180.

Morishita Kiichi. 1984. Shin'ai o arawasu setsubigo 'me' no genshō suru katei ni tsuite: Fukui-shi oyobi sono shūhen chiiki o chūshin ni shite [The decline process of the suffix -me expressing affection: Focussing on Fukui and the neighbouring areas]. *Iwate idai kyōyō-bu kenkyū nenpō* [Iwate Medical University, College of Liberal Arts Research Annual Report] 19. 180-188.

Murasaki Shikibu. ca 1021 [1927]. *Genji-Monogatari dai-2* [The Tale of Genji, vol. 2]. Yosano; Nihon koten zenshū kankō-kai.

Nakamoto Masatomo. 1980. Hachijō-jima hōgen no kenkyū: Goi-teki tokushoku to jintai goi [Study of Hachijō island dialect: Lexical features and the vocabulary of the human body]. *Nihongo kenkyū* [Research on Japanese] 3. 98-111.

Nakamoto Masatomo. 1984. Hachijō-jima [Hachijō island]. In *Tōkyō-to no kotoba* [The languages of Tōkyō Metropole], 157-188. Tōkyō: Tōkyō Metropolitan Board of Education.

Naitō Shigeru. 1979 [2002]. *Hachijō-jima hōgen* [Hachijō-jima dialect]. Tōkyō: Fuchū-shi.

Nitta Tetsuo. 2006. Hōgen ni mirareru ikimono-na ni tsuku setsubiji 'me'- Ishikawa ken Haku Misaki hōgen o chūshin ni [The dialectal -me suffix that attaches to names of living beings]. In Nakai Seiichi & Long, Daniel & Matsuda Kenjirō (eds.), *Nihon no fīrudo gengo-gaku* [Japanese field linguistics], 122-143. Tōkyō: Katsura shobō.

Nomura Denshi & Yanagita Kunio. 1942. *Ōsumi Kimotsuki-gun hōgen-shū* [Dialect collection from Kimotsuki, Ōsumi]. Tōkyō: Chūō kōronsha.

Ōshima Ichirō (ed.). 1986. *Hachijō Kojima hōgen chōsa hōkokusho* [Hachijō Kojima dialect survey report]. Tōkyō: Tōkyō Metropolitan Board of Education, Culture Division.

Ōta Gensai. 1829 [1900]. *Rigen shūran* [Overview of colloquial speech]. Ed. by Murata Ryōa, Inoue Yorikuni, Kondō Heijū (eds.), vol. 3. Tōkyō: Kondō shuppan-bu.

Ōta Nanpo. c. 1811 [1996]. *Ichiwa ichigen* [Small texts with small comments]. Ed. by Nihon Zuihitsu Taisei. Tōkyō: Yoshikawa kōbunkan.

Ōtsuki Fumihiko. 1939. *Daigenkai* [Big sea of words]. Vol. 4. Tōkyō: Tozan-bō.

Pellard, Thomas. 2010. *Ōgami: Éléments de description d'un parler du Sud des Ryukyu*. Paris: EHESS. (Doctoral dissertation.)

Rodrigues, João. 1604. *Arte da lingoa de Iapam composta pello Padre Ioão Rodriguez Portugues da Cõpanhia de Iesv diuidida em tres livros.* Nangasaqui: Collegio de Iapão da Companhia de Iesv.

Sasano Ken (ed). 1956. *Nō kyōgen* [Nō interludes]. Vol. 2. Tōkyō: Iwanami shoten.

Shibata Renzaburō (translator) & Bulwer-Lytton, Edward (author). 1951. *Pompei no saigō no hi* [The last days of Pompei]. Tōkyō: Kaiseisha.

Shimoji Michinori. 2022. Number in the Japonic family. In Acquaviva, Paolo & Daniel, Michael, *Number in the world's languages: A comparative handbook*, 505-528. Berlin/Boston: Mouton de Gruyter.

Shōgakukan. 2001. *Nihon kokugo daijiten* [Dictionary of the Japanese language], vol. 12. Tōkyō: Shōgakukan.

Shōgakutosho. 1989. *Nihon hōgen daijiten* [Dialect dictionary of Japanese].

Suzuki Akira. 1816 [1930]. *Gago onjō-kō* [Essay on the sounds of classical language]. Ed. by Nagoya kokubu-ngaku-kai. Nagoya: Nagoya kokubungaku-kai.

Tachibana Shōichi. 1933. Ushi no hōgen [The dialect forms of 'cow']. *Hōgen* [Dialect] 3(10). 754-775.

Tachibana Shōichi. 1936. *Gengogaku gairon* [Introduction to dialectology]. Tōkyō: Ikuei shoin.

Takahashi Yoichi. 1802 [2012]. En'ō kōgo [Conversations of the garden elders]. In Hachijō-jima no komonjo o yomu kai (ed.), *Hachijō-jima no Komonjoshū* [Collection of old writings from Hachijō island], vol. 1, 1169. Hachijō: Nankai taimusu.

Takemoto Takeshi. 1937 [1976]. *Nagato hōgen-shū* [Dialect collection form Nagato]. Tōkyō: Kokusho kankō kai.

Takiyama Masatarō & Yanagita Kunio. 1944. *Tsushima nanbu hōgen-shū* [Dialect collection from southern Tsushima]. Tōkyō: Chūō kōronsha.

Togawa Naoki. 1966. *Kojima chihō no hōgen-shū* [Dialect collection from the Kojima region]. Kojima: Kojima minzoku kenkyūkai.

Yamada Tatsuya. 1984. Dōbutsumei ni tsuku 'me' ni tsuite: Gifuken Ibi-gun no baai [The *-me* attaching to animal names: The case of Ibi district, Gifu prefecture]. *Nagoya hōgen kenkyūkai kaihō* [Report of the Nagoya Dialect Research Association] 1. 1-6.

Yamada Heiemon. 2010. *Kieteiku shimakotoba: Hachijō-go no keishō to sonzoku wo negatte* [A disappearing insular language: Hoping for the inheritance and survival of Hachijō language]. Tōkyō: Ikuhōsha.

Yamagata hōgen kenkyū-kai. 1970. *Yamagata hōgen jiten* [Dictionary of the dialects of Yamagata]. Yamaga-ta: Yamagata kenkyū-kai.

Yamanaka Rokuhiko. 1967. *Yamaguchi-ken hōgen jiten* [Dictionary of dialects from Yamaguchi prefecture]. Yamaguchi: Yamaguchi-ken chihō shigaku kai.

Central Chadic numerals in comparative perspective

Václav Blažek

Department of Linguistics & Baltic Studies, Masaryk University, Brno
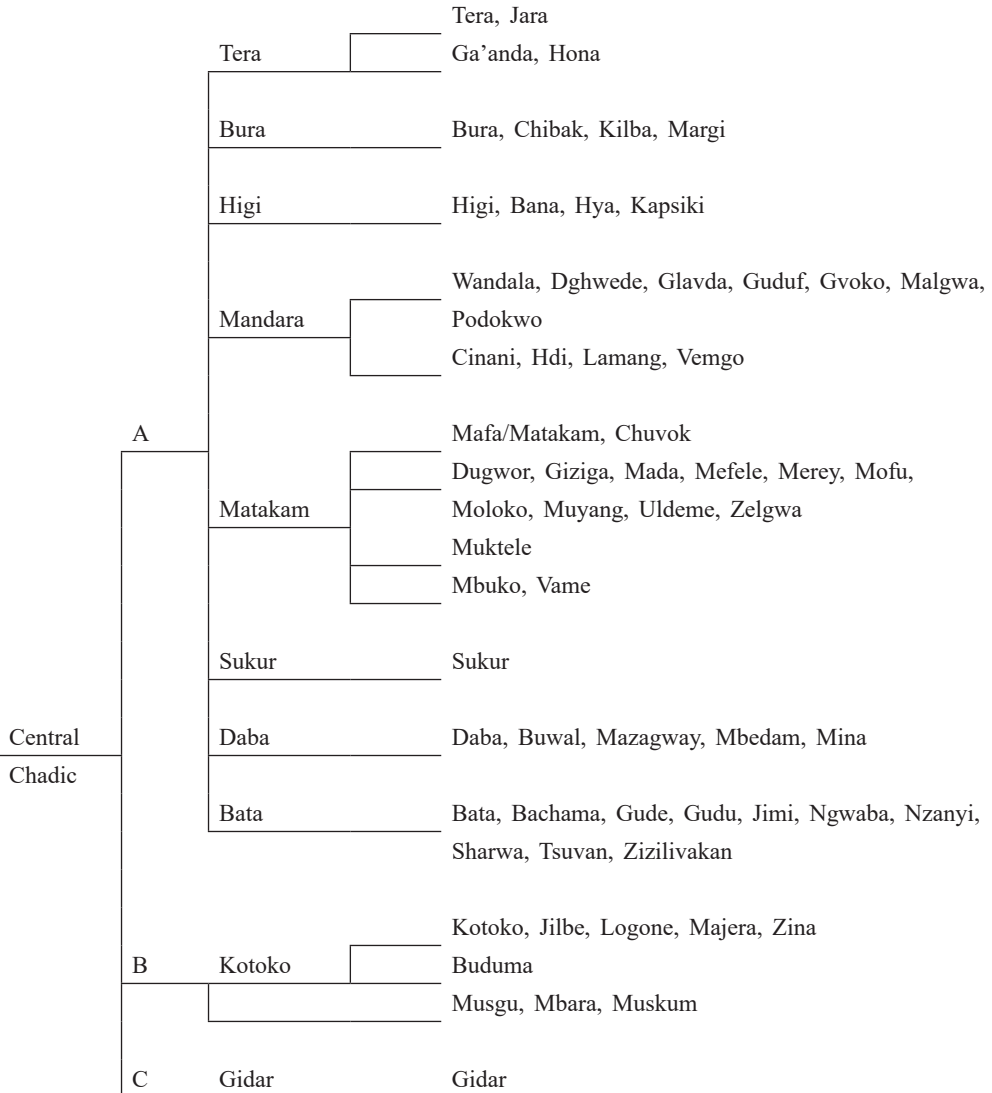
blazek@phil.muni.cz | ORCID: 0000-0002-6797-7188

Abstract: The present study summarizes the forms of numerals of the first decade in more than 200 Central Chadic languages and their varieties, including their sources, to analyze their internal structure and external relations, in both genetic and areal plans.

Keywords: Central Chadic, numerals, internal structure, external relations, etymology

The contribution concentrates on the basic numerals of the first decade in the Central branch of the Chadic language family of the Afroasiatic macrofamily. It was preceded by two studies devoted to West Chadic and East Chadic numerals (Blažek 2018; 2021). The first task is the documentation of all relevant sources, i.e. both the most recent records in scientific transcriptions and historical records of the first travellers from all described languages. The second task is the analysis of their internal structure, verified in their external comparison with corresponding forms in other Chadic languages and finaly the identification of prospective cognates in other Afroasiatic branches or areal parallels in neighboring Niger-Congo or Nilo-Saharan languages. The results are confronted with the most recent variants of classifications of the Central Chadic branch by Barreteau & Jungraithmayr (1993), Starostin (2010), Newman (2013), and Gravina (2011; 2014).

1. Classification of the Central Chadic branch

1.1. The model of classification of the Chadic languages of Paul Newman (2013) is already classical:

```
                                   Tera, Jara
                   Tera ──────────
                                   Ga'anda, Hona

                   Bura ──────────── Bura, Chibak, Kilba, Margi

                   Higi ──────────── Higi, Bana, Hya, Kapsiki

                                   Wandala, Dghwede, Glavda, Guduf, Gvoko, Malgwa,
                   Mandara ───────  Podokwo
                                   Cinani, Hdi, Lamang, Vemgo

          A                        Mafa/Matakam, Chuvok
                                   Dugwor, Giziga, Mada, Mefele, Merey, Mofu,
                   Matakam ──────  Moloko, Muyang, Uldeme, Zelgwa
                                   Muktele
                                   Mbuko, Vame

                   Sukur ──────────── Sukur

Central            Daba ──────────── Daba, Buwal, Mazagway, Mbedam, Mina
Chadic
                   Bata ──────────── Bata, Bachama, Gude, Gudu, Jimi, Ngwaba, Nzanyi,
                                     Sharwa, Tsuvan, Zizilivakan

                                   Kotoko, Jilbe, Logone, Majera, Zina
          B        Kotoko ───────  Buduma
                                   Musgu, Mbara, Muskum

          C        Gidar ──────────── Gidar
```
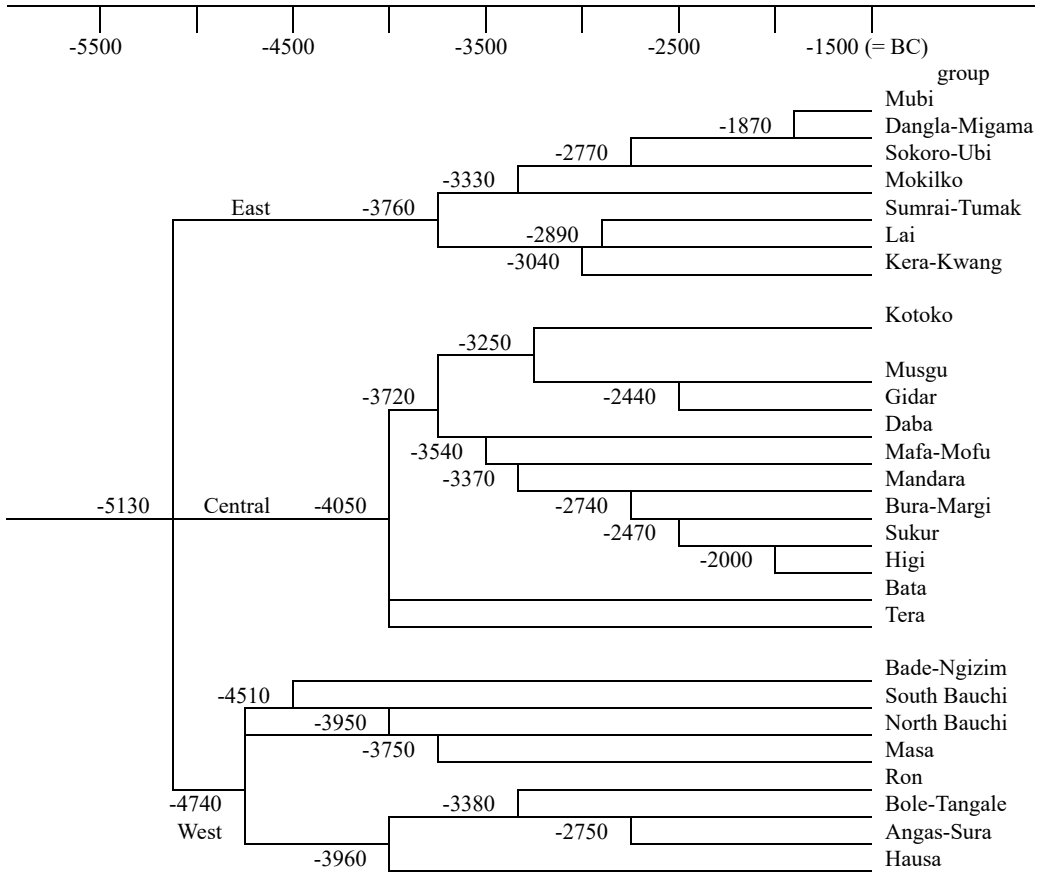
1.2. Richard Gravina (2011: 67-84; 2014: 26-27) limits his research only to the Central Chadic (= Biu-Mandara) branch of the Chadic family. His model is based on his careful analysis of historical phonetics and its changes in perspective of relative chronology. The result is significantly different from the traditional model of Paul Newman:

Central Chadic			
South	Mafa		Mafa, Mefele, Cuvok
	Sukur		Sukur
	Daba	Daba	Daba, Mazagway Hidi
		Mina	Mina, Mbudum
		Buwal	Buwal, Gavar
	Bata	Bata pr.	Bachama, Bata, Fali, Gude, Gudu, Holma, Jimi, Ngwaba, Nzanyi, Sharwa
		Tsuvan	Tsuvan, Zizilivakan
	Tera		Boga, Ga'anda, Hwana
	Hurza		Vame, Mbuko
North	Margi-Mandara-Mofu	Margi	Bura	Bura, Cibak, Kofa, Putai, Nggwahyi
			Margi	Kilba, Margi, South Margi
		Mandara	Wandala	Mandara, (Malgwa)
			Glavda	Cinani, Dghwede, Guduf,
			Podoko	Gava, Glavda, Gvoko Podokwo, Matal
		Tokomberi		Uldeme, Mada, Muyang, Moloko
		Mofu	Meri	Zelgwa, (Gemzek), Merey, Dogwor
			Mofu	North Mofu, Mofu-Gudur
	Higi		Bana, Hya, Psikye, Kamwe
	Lamang		Lamang, Hdi, Mabas
	Marwa		Giziga N+S, Mbazla
	Gidar		Gidar
	Musgum-North Kotoko	Musgum		Mbara, Musgum, Muskum
		North Kotoko		Mpade, Afade, Malgbe, Maltam
		Island Kotoko		Buduma
		Central Kotoko		Lagwan, Mser
		South Kotoko		Zina, Mazera

1.3. Jungraithmayr & Ibriszimow 1994 (JgIb II, xii-xiv) divided the Central Chadic branch into the following groups: Tera, Bura-Margi, Higi, Bata, Lamang, Mandara, Sukur, Mafa-Mofu, Daba, Gidar, Kotoko, Musgu, plus Masa (the strongest arguments for affiliation of Masa into the Central Chadic branch were formulated by Tourneux 1990). But these groups are not parallel, there is an internal hierarchy, which can be illustrated by the following tree-diagram, constructed on the basis of the results published by Barreteau & Jungraithmayr 1993, applying the standard 100-word-list to the mutual comparison:
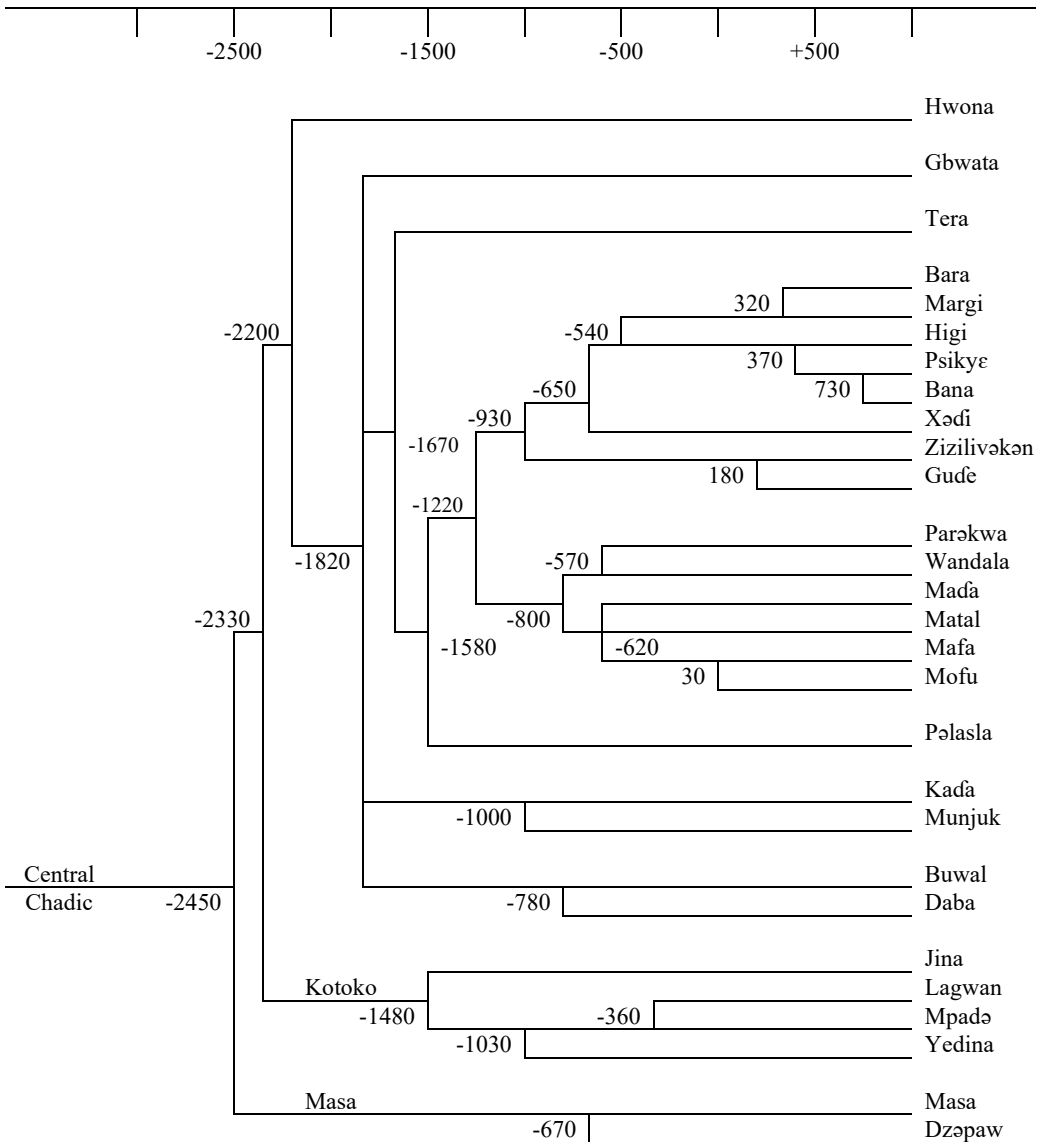
1.4. Another tree-diagram based on the glottochronological test was realized by George Starostin (ms. 2010) for the whole Chadic family. Rather surprising is his inclusion of the group Masa in the West branch, together with North Bauchi.

-5500	-4500	-3500	-2500	-1500 (= BC)	

group

```
                                                          Mubi
                                            -1870         Dangla-Migama
                                  -2770                   Sokoro-Ubi
                          -3330                           Mokilko
      East      -3760                                     Sumrai-Tumak
                                  -2890                   Lai
                          -3040                           Kera-Kwang

                                                          Kotoko

                                  -3250                   Musgu
                  -3720                     -2440         Gidar
                                                          Daba
                          -3540                           Mafa-Mofu
                          -3370                           Mandara
                                  -2740                   Bura-Margi
                                  -2470                   Sukur
                                            -2000         Higi
                                                          Bata
-5130  Central   -4050                                    Tera


                                                          Bade-Ngizim
                  -4510                                   South Bauchi
                          -3950                           North Bauchi
                          -3750                           Masa
                                                          Ron
        -4740             -3380                           Bole-Tangale
        West                       -2750                  Angas-Sura
                  -3960                                   Hausa
```

2. Survey of the Central Chadic numerals

2.1. Tera

Language	1	2	3	4	5
Tera$_{1/2}$	*dà / da*	*rāp / rab*	*kúnúŋ / kunuk*	*vàt / fad*	*qúrmún / ɠurmun*
Tera$_3$	*dà*	*rap*	*kúnú*	*vàt*	*gúrmùn*
Pidlimdi	*ərdə̀*	*rap*	*màkɨn*	*vɜ̀dî*	*gurmun*

Ga´anda	*ařta*	*suřři*	*mahkə̀n*	*fwədà*	*dìrmən*
Boka	*ɍrtà*	*cə̀p*	*məkkən*	*fwədà*	*durmən*
Gaɓin	*hə̀rtà*	*sɍri*	*maxkàn*	*fwədə̂*	*dɨrmə̀n*
Hwona	*tìtal*	*suɣurì*	*maxə̀n*	*fadà*	*tuʃ(ù)*
PTera PHwona	⁺*hiʔarda* ⁺*tital*	⁺*rap* *sɨrri	⁺*kunu* *mahkan	*foɗ = ⁺*fʷaɗ*	*dìrman ⁺*tɨʃᵛ*

Language	6	7	8	9	10
Tera$_{1/2}$	ⁿ*jòŋ / njoŋ*	*mút / mut*	*mʲāsī / miyasi*	*mīʑām / milam*	*gʷàŋ / gwan*
Tera$_3$	*njòŋ*	*mut*	*míyasi*	*mə̀ʑam*	*gwàŋ*
Pidlimdi	*šòŋ*	*mùdî*	*miyérzi*	*mìðam*	*ɓwàn*
Ga´anda	*mɨca*	*mwùt(n)*	*fwətfwədà*	*wə̀nhə̀hə̀ʔařtà*	*kum*
Boka	*tyɛ̀xxɛɫ*	*mwut*	*fwotfwə̀dà*	*hàhìrta*	*kum*
Gaɓin	*mɨcè*	*mwòt*	*fwətfwədà*	*fendɨrmən*	*kùm*
Hwona	*mìki*	*mɨd(u)*	*(w)ùvwədà*	*wùtàrè*	*gumdìdî / kum*
PTera PHwona	⁺*ŋyon* ⁺*miki*	*muɗ = ⁺*mɨdu*	⁺*miyasɨri* ⁺ *fʷadfʷaɗ*	⁺*miʑam* ⁺*gwaŋ-ha-hiʔarda* ⁺*fʷaɗ dìrman*	⁺*gwaŋ* *kʷɨm

Boka = Boga numerals by Charles H. Kraft https://mpi-lingweb.shh.mpg.de/numeral/Boga.htm
Ga´anda numerals by Charles H. Kraft 2007 https://mpi-lingweb.shh.mpg.de/numeral/Gaanda.htm
Gaɓin numerals by Charles H. Kraft (1981, II: 33)
Hwona numerals by Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/Hwana.htm
Pidlimdi numerals by Charles H. Kraft (1981, II: 4)
Tera$_{1/2}$ numerals by Babayo Madi (2014) / Dr. Musa (1995) https://mpi-lingweb.shh.mpg.de/numeral/Tera.htm
Tera$_3$ numerals by Paul Newman (1964: 36).

2.2. Bura – Margi

Language	1	2	3	4	5
Bura$_1$	*ntang*	*suda*	*makər*	*nfwar*	*ntufu*
Bura$_2$	*ntàŋ*	*sùdà*	*mākə̀r*	*mfwar*	*ntə̀fù*
Bura-Pabir$_{1/2}$	*ntàng / ntaŋ*	*sùdà / sudà*	*màkə̀r/makùř*	*fwàr/fwař*	*ntìfù / ntufù*
Chibak	*tə̀ŋ / patù / dukù*	*sudɛ̀*	*makùř*	*fwòdu*	*tufù*
Ngwaxi$_1$	*tə̀ŋ*	*sɨdà*	*makùř̃*	*fwə̀r̃*	*tufù*
West Margi	*duku / təŋ / dugu*	*sudà / fɨdɛ̀*	*makùr*	*fɔdu/fwodu*	*tufù*
Central Margi	*taŋ / pału / tɨtikù*	*mɨłù / sɨdàŋ*	*makùř̃*	*fwodù*	*ntɨfù*

South Margi	*tà'ú, zə̀mù, páthlú*	*sə̀dà'u / mə́thlú*	*màkə̀r*	*fwádú/ fódú*	*tə̀ffú*
Hildi	*ʒaŋuw / tàyu*	*míɬū / sɨ̀dàyū*	*mákr̃ù/makùr̃(ù)*	*fwɔ̄dù*	*tufù*
Wamdiu Margi	*zìmu/tə̀ŋu/paɫu*	*miɫu*	*makɨru*	*fodu*	*tufu*
Kilba₁	*dzàŋ*	*mətlù*	*màkə̀r/màkərù*	*fòdù*	*tùfù*
Kilba₂	*žáàŋú*	*mέlù*	*mákrù*	*fôdù*	*tûfû*
*	*⁺taŋ / ⁺tiku / ⁺duku / ⁺dzaŋaw / ⁺paɫu*	**suda = ⁺sɨ́dayaw ⁺miɫu*	**maakir = ⁺maakiru*	**fʷadu*	**tifʷ*

Language	6	7	8	9	10
Bura₁	*nkwa*	*murfa*	*ncisu*	*umdla*	*kum(a)*
Bura₂	*nkwà*	*murfà*	*ncìsù*	*umdlà*	*kuma*
Bura-Pabir₁/₂	*nkwà / ŋ̀kwà*	*mùrfà / murfà*	*cìsù / ncɨsù*	*ùmðlà / ḿðà*	*kùmà / kuma*
Chibak	*ŋ̀kwà*	*murifwὲ*	*ntsisù*	*miðæ*	*kuma*
Ngwaxi₁/₂	*nkwɔ̀*	*mur̃fà*	*ncis / ncis mɨðà*	*mɨða / mɨða*	*kuma*
West Marghi	*kwa / kwɔ̀*	*mudúfā / mudîfὲ*	*cisù / ncɪsù*	*ḿðà / mðὲ*	*kuma / kumɛ*
Central Marghi	*ŋ̀kwà*	*mɪdífù*	*ntsisù*	*ḿðù*	*kumu*
South Margi	*kwà*	*mὲ̀dὲ̀fàù*	*císsú*	*ə̀ldlàù*	*kùmòù*
Hildi	*kwà*	*mídɨ́fɔ̄*	*císū*	*ðə̄w*	*kúmɔ̀*
Wamdiu Margi	*kwà*	*mɨdufə̀w*	*cisù*	*ɨ̀ðə̀w*	*kumò*
Kilba₁	*kwà*	*mədəfà*	*cìsù*	*dlà*	*kùmà / kùm*
Kilba₂	*kwà*	*mē̄dɨ̄fà*	*čīsù*	*ðà*	*kúmà*
*	**kʷa = ⁺ŋkʷa*	**mɨdɨfaw*	**tsʲisɨw*	*⁺miɫaw*	**kʷɨma*

Bura₁ by Roger Blench (2010).
Bura₂ numerals by Russell G. Schuh with Elisha Shalanguwa (1981-82).
Bura-Pabir₁/₂ numerals by Mohammed Aminu Muazu and Mr. Fibi Balami (2010) / Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/Bura.htm
Central Marghi numerals by Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/ Marghi-Central.htm
Chibak = Kibaku numerals by Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/Cibak.htm
Hildi numerals by Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/South%20Marghi. html
Kilba₁ = Huba numerals by Mohammed Aminu Muazu (2010)
https://mpi-lingweb.shh.mpg.de/numeral/Huba-Kilba.htm
Kilba₂ numerals by Charles H. Kraft (1981, II: 92)
Ngwaxi₁ = Nggwahyi numerals by Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/ Nggwahyi.htm
Ngwaxi₂ = Nggwahyi numerals by Charles H. Kraft (1981, II: 82)
South Marghi numerals by Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/South%20 Marghi.html
Wamdiu Margi numerals by Charles H. Kraft (1981, II: 112)
West Margi = Putai numerals by Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/ Putai.htm

2.3. Bana

Llanguage	1	2	3	4	5
Bana₁/₂	*tánə̀ / kwətiŋ*	*bákə̀ / bakə*	*máhə̀kánə̀ / mahkan*	*fádə̀ / fadə*	*cífə̀ / cifə*
Fali Kiria₁	*gutàn / tanəy*	*ɓwukuʔ*	*màkun(u)*	*fwadû̃ʔ*	*(w)cɪfuʔ*
Fali Kiria₂	*gùtàn*	*ɓə̀kə̀*	*makən*	*fwàdə̀ / fwàr*	*ncɪfə*
Baza	*gutá*	*bagé*	*makená*	*fõra*	*jufa*
Fali Wuba	*tan*	*baka*	*makin*	*nfwo̧r*	*ncifa*
Fali Gili	*kwùtyɪn / taŋ*	*bak*	*màxkan*	*fwaɗ*	*cif*
Fali Mucella	*tèn, ʔaȓ̃mə*	*bek / buk*	*màxk(u)*	*fwəɗ*	*tuf*
Fali ɓwagira	*tân*	*buk*	*màxk(un)*	*fwat*	*tuf*
Higi Ghye	*paðɛ / tanɛ*	*ɓagɛ*	*màŋkɛ*	*fwadɛ*	*wcivi*
Kapsiki₁	*kwetɛŋe*	*bake*	*mahekene*	*wəfade*	*mcɛfe*
Kapsiki₂	*kwə̀téné*	*bákə́*	*màRkə́n*	*wə̀fádə́*	*ùcəfə́*
Kapsiki₃	*kōtán*	*bǎk*	*mǎchkə́n*	*ōfǎ́t*	*nschǎf*
Hidji	*kūtə́*	*bāgə́*	*mākə̆nə́*	*foārə́*	*tschifə́*
Higi Kamale	*tanɛ, kùtanɛ*	*bakɛ*	*màxkunɛ*	*fwadɛ*	*mcufwɛ*
Higi Nkafa	*paɬɛ / kùte*	*ɓwàge*	*màkìne*	*fwàre*	*ncìfe*
Higi Baza	*paɬəw*	*ɓwògu*	*màkə*	*fwô*	*mcìfe*
Higi Futu	*tɪxuy*	*ɓaxu / ɓaku*	*màku*	*fwadu*	*mcɛf / mcaf(i)*
Fali Jilbu	*lɪm*	*sul*	*màxku*	*fwəy*	*mùxtyup*
Zizilivakan	*lɪm*	*sul*	*màxku*	*fwəy*	*mùxtyup*
*	⁺*tɪx* / ⁺*tanay* / ⁺*paɬaw* / ⁺*lim*	*ɓɪwak = ⁺*ɓɪwaku* / ⁺*sul*	*maxkɪn	*wifadî = ⁺*fwadî*	*wɪtsʲifɪ ⁺*mɪtsʲifɪ*

Language	6	7	8	9	10
Bana₁/₂	*kwáŋ / kwaŋ*	*bə̀rfàŋ / mbə̀rfəŋ*	*də̀ɣə̀sə̀ / dəghəs*	*mə̀lísdə̀ / məsliɗ*	*mə̀ŋ / məŋ*
Fali Kiria₁	*ŋkwaŋ*	*mbùrùfũŋ*	*tùɣùsùʔ*	*ǹwɬti(ʔyì)*	*gwùm(ù)*
Fali Kiria₂	*kwáng*	*mbìrfə́ng*	*tə́ghə̀s*	*wr̩ì*	*gùm*
Baza	*kwānge*	*brfunga*	*tikisa*	*thi*	*mūnge*
Fali Wuba	*nkwaŋ*	*birfuŋ*	*tɪxis*	*mti*	*gum*
Fali Gili	*kwəŋ*	*mbùrùfuŋ*	*dùɣùs*	*mɪɬtɪt*	*mùŋ*
Fali Mucella	*yidə̀w* < PFali *yidø	*mbùr̃fuŋ*	*tùɣus*	*mìðɪŋ*	*gùm*
Fali ɓwagira	*mbùrùfuŋ kuwà*	*mɪdûf*	*tùɣus*	*mìðî̃ŋ*	*po gumu*
Higi Ghye	*kwaŋəy*	*mbùr̃ùfəŋəy*	*tùgùzi*	*wɪɬti*	*mùŋəy*

Kapsiki₁	ŋkwaŋe	mberefaŋe	deghese	mesli	meŋe
Kapsiki₂	ùŋkwáŋə́	mbrə̀fáŋə́	dyə̀ssə́	mìєí	mə̀ŋə́
Kapsiki₃	nkŏng	bŭrfắn	dĕrss	michtí / miltí	măng
Hidji	nkōắ	mbĕrfŭ́ngĕ	tĕgēssĕ́	tī́	mĕngĕ́
Higi Kamale	ŋkwaŋəy	mbùrùfaŋəy	də̀ɣə̀sε	mìlti	mùŋε
Higi Nkafa	kwaŋəy	mbùr̃fuŋəy	tìkìse	pɬtìyi	mùŋəy
Higi Baza	kwa	mə̀lùfə	tə̀ɣə̀sə	pɬtì	mùŋə
Higi Futu	kwaŋ(u)	mbùr̃fəŋ(u)	dìɣùsu	pɬtìyi	mùŋi
Fali Jilbu	ŋ̀kwaʔ	mbùrfìŋ	tə̀ɣìs	mìðî	gumù
Zizilivakan	ŋ̀kwaʔ	mbùrfìŋ	tə̀ɣìs	mìðì	gumù
*	*kʷaŋ = ⁺ŋkʷaŋ	*ᵐbirfìŋ	*tɨɣis	⁺miʐi	*gʷɨm

Bana₁/₂ numerals by Ndokobaï (2009 / Zra Simon and Robert Hedinger (1994
https://mpi-lingweb.shh.mpg.de/numeral/Bana.htm
Baza numerals by Johannes Lukas (1937: 113).
Fali ɓwagira numerals by Charles Kraft (1981, III: 26).
Fali Kiria₁ = Kwame numerals by Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/Kwame.htm
Fali Kiria₂ numerals by Roger Blench
http://www.rogerblench.info/Language/Afroasiatic/Chadic/Central/Bura/Bura%20opening%20page.htm
Fali Gili numerals by Charles Kraft (1981, II: 191).
Fali Jilbu = Zizilivakan numerals by Charles Kraft (1981, III: 6).
Fali Mucella numerals by Charles Kraft (1981, III: 16).
Fali Wuba numerals by Johannes Lukas (1937: 113).
Hidji numerals by Hauptmann Strümpell (1923: 120-23).
Higi Baza numerals by Charles Kraft (1981, II: 141).
Higi Futu numerals by Charles Kraft (1981, II: 171).
Higi Ghye = Hya numerals by Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/Hya.htm
Higi Kamale numerals by Charles Kraft (1981, II: 151).
Higi Nkafa numerals by Charles Kraft (1981, II: 131).
Kapsiki₁ = Psikye numerals by Hubert Nkoumou (2007) https://mpi-lingweb.shh.mpg.de/numeral/Psikye.htm
Kapsiki₂ numerals by Véronique de Colombel (letter from the Oct 20, 1987).
Kapsiki₃ numerals by Hauptmann Strümpell (1923, 120-23).
Zizilivakan numerals by Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/Zizilivakan.htm

2.4. Lamang – Mandara

Language	1	2	3	4	5
Lamang	tíuwá / tálá	χésá	χkə́ná	ùfádă	χʷtáfá
Alataghwa	tala	xasha	xəkŋa	úfada	xutáfa
Turu	teku, nden	xes, ghis	xakn	ufat	xutaf
Vemgo/Vizik	ndıŋ, tala	xrεs	xkın	ufat	xtaf
Vemgo-Mabas	pál / tékw	hés	xə̀kə̀n	úfáɗ	xútáf
Gvoko	palò / tekò	xecò	xək̄ʷarò	fwadò	ʐaʔò
Hdi	tèkw	hìs	hə̀kə̀n	fwáɗ	hùtáf

Hidi	tɘrtúk	Rís	Rɘkán	fɘ́wáɗ	kɘ̀táf
Dghwede	titikwì, tekwè	micè	xɘkùrè	fidì	ðiɓi
Gava	cɘk/kyek, kìtɘkwùlà	mits(à)	xìkur̃d(à), x̀kur̃	ʔùfɘd(à)	ðiɓ(à)
Guduf-Gava	tekʷè / kitakʷè	mitsè	xɘkɘrdè	ùfɘdè	ɮ̀ɓè
Cinene	pàlà	bùʷà	xɘ̀kɘ̀rdà	ùfàdã	ɮ̀ɓà
Glavda₁	páll	bwa	xkɘrɗ	ufáɗ	ɮɘɓ
Glavda₂	pala, kika	bʊwà	x̀kʊrdà	ʔufadà	ðiɓà
Glavda₃	pállà, cíka	bù(a)	xkɘ́rdà	úfàd(a)	dlɘ́ɓâ
Nakatsa	pala	buwa	xakarda	wòfada	lìba
Mora	pállé	bɘ̀wà	kɘ́dʒ̃́yá	ùfàdè	íɮ̀ibè
Wandala₁	pálle	búwa	kɘɟyé	ufáde	iiɮɘbé
Wandala₂	pɘle	(m)bùwà	kiɟye	ʔùfwàdè	ʔiðìbè
Wandala₃	pālɘ́	bŭĕ	kīdɘ́	ōfádè	īldebe
Wandala₄	páĺle	bwa	kɘ́ɟẹ	ufade	ildebe
Wandala₅	palle	búā	keyē	úfadē	íldebē
Wandala₆	mtɘ̆kwɘ́	bùà	kí'jé	ĩ́ʒ̃̀bè	ĩ́ʒ̃̀bè
Wandala₇	mtaqué	sardah	kighah	fuddah	eliba
Podokwo₁	kutɘra	sɘra	makɘra	ufaɗa	zlama
Podokwo₂	boktera	ezera	makera	ofada	zama
Podokwo₃	kwɘ̄trá	sɘ̀rà	màkɘ̀rà	ùfàdà	ɮ̀màà
PLamang	⁺palaw // ⁺tikʷɨ // ⁺tala // ⁺dɨn	⁺hisɨra	*hikɨna	*wifaɗ	*hʷitaf
PMandara	⁺palaw // ⁺tikʷɨ	*bɨwa // ⁺mitsay	*hikɨri = ⁺hikɨrɗa	*ufadî	*ɮidîm = ⁺ʔiɮiɓa
PPodoko	⁺kʷitɨra	*isɨra	⁺makɨra	⁺ufaɗa	⁺ɮama

Language	6	7	8	9	10
Lamang	m̀kwá / m̀kuwá	ɘ̀lfáŋá	tɘ̀ɣásá	tɘ̀mbáyá	ɣwáŋá
Alataghwa	ŋkua	lfaŋga	tghasa	tɘmbáya	ghuaŋa
Turu	muku	ndfan	tɘghas	?	ghwan
Vemgo/Vizik	ɔŋko, mku	rɘfaŋ	tɪxas	tɪmbai	(ɔ)waŋ
Vemgo-Mabas	ńku	lɘ̀fàŋ	tɘ̀ɣàs	tɘ̀mbàj	yɘ̀wàŋ
Gvoko	ŋkoyò	ntfaŋgò	tɘ̀ɣasò	tìmbayò	ɣʷaŋgò
Hdi	mɘ̀kúʔ	ndɘ̀fáŋ	tɘ̀ɣás	tɘ̀mbáy / timbe	ɣwàŋ
Hidi	mɘ̀kwɘ́	ndɘ̀fáŋ	tɘRás	tìmbéɗ	Rɘ̀wáŋ
Dghwede	ńkwe	wudîfi	tɘɣɘ̀še / tɘxɘse	tɘmbɘ̀	ɣwàŋga
Gava	ŋ̀kwaxà	wùdîf(à)	tɘ̀ɣɘs(à)	vɘsulɘ̀mbɘ̀d(à)	kɘldûk(a)
Guduf-Gava	ŋkʷaxè	ùèdîf	tɘ̀ɣɘsè	vaslambàdè	kulɘkè
Cinene	ŋkʷàxà	ùdîfà	tɘ̀ɣsà	vaslambàdà	klawà

Glavda₁	ŋkwax	udîf	tə́xs	vaslambaɗ	klàáwá
Glavda₂	ŋ́kwàxà	wudîfà	təxsà	vàsòlə̀mbàdà	kʊlàwà
Glavda₃	ǹkwax(a)	údîfà	tə́xsà	vàsəlambadà	kə̀lawa
Nakatsa	ǹkwaxà	wùdîfè	təɣəsə	vəsə̀lə̀mbə̀da	kalawò
Mora	ə́ŋkwáhé	və́wyè	tísè	másə́lmànè	kə̀làwà, kəladʒubəwa
Wandala₁	unkwé	vúye	tiise	másə́lmane	kə́láwa
Wandala₂	ǹkwàxɛ̀	vuyè	tisè	məsilmànè	kùlàwà
Wandala₃	nkŏhě	tīsě	măsélmānē	tschémtuke	klǎ̆uĕ
Wandala₄	uŋkohe	vúyẹ	tise	masilmane	išumi
Wandala₅	úŋkohē	wūye	tīse	masilmánne	kelaua
Wandala₆	úŋkwẹ́hẹ́	vǘyè	tísè	mà̰sélmànè	kḛ̀làwà
Wandala₇	n'quaha	vanyah	tísah	musselman	klaaň
Podokwo₁	məkuwa	madəfa	za	metirəce	jima
Podokwo₂	mokūa	madufa	eza	metecar	χima
Podokwo₃	mə̄kwá	mādə̄fá	ə̄zzá	mātə̄rcá	dʒìmìyà, dʒim dar səra
PLamang PMandara PPodokwo	*mkɨwa *mɨkɨwah ⁺mɨkɨwa	⁺rɑ̆faŋ *mɨdîfɨ ⁺vabɨwa ⁺madɨfa	*tɨɣis *tɨɣis ⁺izza < ⁺isira dʒim 10-2?	⁺timbaɗay ⁺visulimbaɗay ⁺masilmanay ⁺matirtsay	*ɣʷaŋ *kɨlawa ⁺dʒim

Alataghwa numerals by H. Büchner via Ekkerhard Wolff (1971: 64).

Cinene numerals by Kim Hak-Soo (2008) https://mpi-lingweb.shh.mpg.de/numeral/Cinene.htm

Dghwede = Zəɣwana numerals by Charles Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/Dghwede.htm

Gava numerals by Charles Kraft (1981, III: 123).

Glavda₁ numerals by Jonathan Owens (2007) https://mpi-lingweb.shh.mpg.de/numeral/Glavda.htm

Glavda₂ numerals by Charles Kraft (1981, III: 98).

Glavda₃ numerals by E.L. Rapp apud Ekkerhard Wolff (1971: 68).

Guduf-Gava numerals by Kim Hak-Soo (2008) https://mpi-lingweb.shh.mpg.de/numeral/Guduf-Gava.htm

Gvoko numerals by Kim Hak-Soo (2008) https://mpi-lingweb.shh.mpg.de/numeral/Gvoko.htm

Hdi = Hdi numerals by Cindy Langermann (1994) + Ndokobaï (2008) https://mpi-lingweb.shh.mpg.de/numeral/Hadi.htm

Hidi numerals by Véronique de Colombel (letter from the Oct 20, 1987).

Lamang = Hiɗkala numerals by Ekkehard Wolff 1988 https://mpi-lingweb.shh.mpg.de/numeral/Lamang.htm

Mora numerals by Véronique de Colombel (letter from the Oct 20, 1987).

Nakatsa numerals by Charles Kraft (1981, III: 135).

Podokwo₁ numerals by Elizabeth Javis https://mpi-lingweb.shh.mpg.de/numeral/Podoko.htm

Podokwo₂ numerals by Johannes Lukas (1937: 127).

Podokwo₃ numerals by Véronique de Colombel (letter from the Oct 20, 1987).

Turu numerals by P.K. Eguchi via Ekkerhard Wolff (1971: 64).

Vemgo-Mabas numerals by Ndokobaï (2008) https://mpi-lingweb.shh.mpg.de/numeral/Vemgo-Mabas.htm

Vemgo/Vizik numerals by Charles K. Meek apud Ekkerhard Wolff (1971: 64).

Wandala₁ (Malgwa) numerals by Doris Löhr https://mpi-lingweb.shh.mpg.de/numeral/Wandala.htm

Wandala₂ numerals by Charles Kraft (1981, III: 86).

Wandala₃ numerals by Kurt Strümpell (1923)
Wandala₄ numerals by Jean Mouchet (1950: 58-59).
Wandala₅ numerals by Heinrich Barth (1862: 9, 11).
Wandala₆ numerals by Johannes Lukas (1937: 121).
Wandala₇ numerals by Julius H. Klaproth (1862: 34-36).

2.5. Mafa – Mada

Language	1	2	3	4	5
Muktele₁	tékwə́lá	sə̀là	mákr	ùfàɗ	ǥii
Matal	dì / tēkùlā	sīlà	màkìr	ùfàɗ	ìǥù
Giziga₁ N.	ɓlà, tiɗka, pàl	cêw	mà:kàr	m̀fàɗ	ǥòm
Giziga₂ S.	plá	cúw	máakàr	mə̀fáɗ	ǥúm
Giziga₃	ɓúlâ, ɓə́lâ	cêw	màakàr	mùfàɗ	dlòm
Giziga₄	sala	tschu	măkĕr	mfắt	sŏm
Muturua	plola	tschiu	makir	mufát	sūin
Muktele₂	den	tschu	mākắng	mo(a)fod	lindam
Baldamu	ɓlà card., ndéŋ det.	cáw	màakáŋ	mòovún/l	ziíndám
Zelgwa₁	īlék	sə́lā	màkár	fāɗ	èǥām
Zelgwa₂	ilík	súla	màkər	əfáɗ	ə̀zləm
Zelgwa₃	lęk	súlạ	mákạr	fạd	ẑạm
Matakam₂	lek	sŭlŏ	măkắr	fũắt	lam
Dugwor	bek	səla	makar	məfaɗ	zlam
Matakam₃	stạd	čęčew	mákạr	fạd	ẑạm
Magumaz	sə̀tàɗ	cawə	máhkàr	fàɗ	ǥám
Mafa₁	sə́táɗ	cew / cecew	makár	fáɗ	zlám
Mafa₂	stàɗ	càw	màkâar	faɗ	ðam
Mafa₃	sə́tá	cawə	màkár	fāɗ	ǥám
Sulede	stá	cawə	màkàr	fàɗ	ǥám
Mufo	sta	tscho	măkắr	fat	ldam
Merey	nə̀tê	súlò	màkàr	fàɗ	ǥàm
Mikiri	ǹtáy	sɨlá	màkàr	mùfàɗ	ǥɔ́m
Mofu N.	nettey	suho	makar	fáɗ	ǥàm
Mofu₁	áçtạ	čew	mākạr	fwọt	dlọm
Cuvok	ámə̀tà	át͡ʃew	máákàr	fáɗ	ǥám
Matakam₁	umtă	tschischéo	mackkar	fũắt	lắm / lsám
Mefele₁	mə̀tá	cécèw	màhkár	fwàɗ	ǥàm
Mefele₂	mə̄tá	cecewə	màhkár	fəwàɗ	ǥám
Mada₁	fə̄tēk	sə̄lá	māhkār	ūfáɗ	ǥàm

Mada₂	ftek	séla	mahkar	wfaàë	zzlaèm
Mada₃	fətɛ́k	sạlạ	mạ̣ḥkər	fạd	ẑạm
Muyang₁	bílīŋ	cìyâw	màhkār	fāɗ	ɮàm
Muyang₂	bílìŋ	tʃŷ	màhkə̄r	fāɗ	ɮàm
Moloko	bìléŋ	tʃɛ̃w	màkár	ùfáɗ / mɔ̀fáɗ	ɮɔ̀m
Baka	bə̀lèŋ	cew	màkar	wúfaɗ	zlàm
Uldeme₁	fɛ̄lɛ́ŋ	brɛ̄tʃâw / tʃâw	mākár	mə̄fáɗ	ɮàm
Uldeme₂	sēléŋ	(bə̀rə̀)céwə́	màkàr	mə̄fáɗ	ɮàm
Uzlam	šélẹŋ	brẹ̄čọ	mạ̄kạr	mfad	ẑạm
Mofu Gudur	teɗ / ték (counting) pál (enumeration)	t͡sew	máakar	məfaɗ	ɮam
Mofu₂	pal / tek	cew	máakar	məfaɗ	ɮam
Mufu	kĕrtĕk	tschéū	makaŋ	fŭdő	dĕrő́m
Mbuko₁	kərtek	tsew	maakaŋ	fudo	ɗara
Mbuko₂	kə́rtek	cew	máákan	fə́do	d̆ara
Mbuko₃	kírtẹk	čẹ́u	mạ̄kạn	fúdọ	dárạ
Hurzo₁	ɓīlē	ciyaw	máŋgān	fə́wdāw	d̆árā
Hurzo₂	bílẹ	čọ̄	mákạn, māŋgạn	fŭ́do	dárạ
Vame	ɓìlɛ́	tʃâw	máŋgàn	fú:dâw	dá̆:rà
PM-Mᴿ PMuktele PZulgo PMafaᴳ PMefele PMada PMofuᴳ PMbukoᴳ PHurzo PMarwaᴳ PUldeme	*bili(n) ⁺tɨkʷ-ila ⁺ʔilik ⁺sitad ⁺(ʔa)mita ⁺fɨ-tik ⁺tikʷ// *pal ⁺-tɨkʷ ⁺ɓilay ⁺pal/⁺ɓila/⁺sala/ ⁺den ⁺šeleŋ	*səl / *tsaw ⁺sɨla ⁺sɨlaw ⁺tsaw ⁺(ts)atsaw ⁺sɨla *sɨwla *tsaw ⁺tsiyaw *tsɨw ⁺bɨri-tsɨw	*ma-kar ⁺makɨr ⁺makɨr *mahkar *mahkar *mahkar *mahkɨr *maakan ⁺máŋgan *maakaŋ ⁺makar	*ma-faɗ ⁺ʔufaɗ ⁺ʔifaɗ *faɗ ⁺fɨwaɗ ⁺wɨfaɗ *wifaɗ *fuɗaw ⁺fuwɗaw *mufaɗ ⁺mifaɗ	*dlam ⁺ɮiw ⁺ʔiɮam *ɮam ⁺ɮam ⁺ɮam *ɮim ⁺dara(m) ⁺dara *ɮiⁿdam ⁺ɮam

Language	6	7	8	9	10
Muktele₁	mūkwā	mə̄də̆f	mìtìgìš	làdgà	kùlà
Matal	mùkʷā	mìdɨf	m̀tìgìʃ	làdìgà	kùlù
Giziga₁ N.	mérkêɗ	tà:rnà	dà:gàfàd	ngòltêr	krô
Giziga₂ S.	mérkéɗ	tàrnà	dàaŋàfáɗ	ngòltír	kúrú
Giziga₃	méerkêd	tàarnà	dàagàfàd	ŋò(o)ltîr	kúrû
Giziga₄	mĕrkín	tārnă̆	dagafat	woltir	kŭrǔ́
Muturua	maerké	tārnă	dǎ̆ngă̆fã	gōltĭr	kŭrrŭ

Muktele$_2$	mĕ́rkĕ	tŭ́rnū	sābĭr	sabla	kŭ́rū
Baldamu	mérkì	tùrú	sàabùr	zàá.ɓə̀là	kúrú
Zelgwa$_1$	ndílék	təsəla	camakar	cə́ɗ	ká̇ró
Zelgwa$_2$	ndílík	təsəlá	tsàmàkə̀r	tswíɗ	kúrwá
Zelgwa$_3$	ndílę̣k	tə́səlạ	tsạmáḳạr	čwǫ̈d	kúrạ
Matakam$_2$	dĕlélék	dĕsĕlắ	samaka	schŏt	kŭrő̌
Dugwor	mukwa	tsela	tsaamakar	tseuɗ	kurow
Matakam$_3$	mó̧ḳọ	tsárạd	tsạmáḳạt	čüd	kúlạ
Magumaz	mə̀kwá	càrạɗ	càmàkàd	cəwéd	kwə̀là
Mafa$_1$	mokwa	tsárad	tsamakaɗ	cə́ɛd	kula
Mafa$_2$	mòkwâ	tsarad	tsə̀məkàd	cud	k(w)ụła
Mafa$_3$	mə̀kwá	cárə̇ɗ	càmàkàd	cə́wéd	kwə̀là
Sulede	mə̇́ká	cárə́ɗ	càmàkàd	cíɗ	kwə́là
Mufo	mŏkbắ	sĕrrắ	zămăkắt	tschoet	pella
Merey	m̀kô	tàsə́là	tsà:mà:kàr	cǒ̃ɗ	krôw
Mikiri	mùkʷó	tə́sə́là	tsámàkàr	cǽd	kúrró
Mofu North	mukó	taasə́lá	tsamakàŋ	tsə́ɗ	kuro
Mofu$_1$	múkọ	číla, tsə́la	čạ́kər	tsúfwạt	kúrọ
Cuvok	máákwà	tásə̀là	t͡ʃǎákàr	t͡ʃ́üɗ	kùràw
Matakam$_1$	mūkŏắ	zēlát	tzắchkăr	tschūĕt	dumbauk
Mefele$_1$	mòkwá	tsə̀làd	t͡ʃ́ǎhkàr	t͡ʃ́üɗ	dùmbók
Mefele$_2$	mə̄kwá	tə̀sə̀làd	cāhkàr	cə̄wd	də́mbáku
Mada$_1$	mə̄kwó	ɤāslá	ɤālāhkár	ábùlùmɓò	dzùmòk
Mada$_2$	mokkoà	slaasélaà	slalahkaàr	oàɓoèlmɓoè	dzmoèkw
Mada$_3$	múkwǭ	ɤạtsạlạ	ç̣álạkạr	abúlumɓọ	dzúmọk
Muyang$_1$	mùkwō	ādə́skə̄lā	āǥálárkār	āmbə́lmɓō	kə̄rāw
Muyang$_2$	mòkʷū	ādə́skə̄lā	āǥáláxkər	āmbólmɓō	krū
Moloko	mòkʷɔ̀	fìse̋re̋	łálákár	hólómɓ́ɔ	kʷɔ̀rɔ́
Baka	múku	sisíri	slalákúr	holombo	kru
Uldeme$_1$	mōkō	sə̄sə̄lā	fə̄rfáḍ	álbìt	kōlō
Uldeme$_2$	mə̄kwə̇́	sə́sə́lā	fə́rfàḍ	élbìt	kə̄lāw
Uzlam	mŭku	sə́səla	fə́rfạd	ạlbit	kúlọ
Mofu Gudur	maakwáw	maasála	daaŋgafaḍ	ʒam-leték / -leteḍ	kúráw
Mofu$_2$	maakwáw	maasála	daakafaḍ	ʒam ləték	kúráw
Mufu	b(ĕ)rkắ	tschibe	schumakang	dĕsūdŏ	kru
Mbuko$_1$	mbə̀rka	tsuwɓe	dzə̀maakaŋ	də̀sudo	kuro
Mbuko$_2$	mbə́rka	cúɓe	jə̀máákan	dɔ̀sə̇́do	kúró
Mbuko$_3$	mbúrkạ	číɓę̣	nə̀fúdọfudọ jumą̣kən	dúsudọ	kúrọ
Hurzo$_1$	márkā	cíɓā	ʒìrè	tóhkē	dʒèm

Hurzo₂	márką	číbą	žiŕę	tąḥkę	ǰim
Vame	márkà	tʃíbà	ʒíːrɛ̀	táhkè	dʒɛm
PM-Mᴿ	*muku		*tlal-kar	*-lumbo	*kuraw / *dzim
PMuktele	⁺mukʷa	⁺mɨdɨ̂f	⁺mitigiš	⁺ladɨga	⁺kulaw
PZelgwa	⁺ndilik	⁺tɨsila	⁺tsamakɨr	⁺tsuwɨd	⁺kuraw
PMafaᴳ	*makʷa	*tsarad	⁺tsamakard	⁺tsuwad	*kʷɨrawʷ
PMefele	⁺mikʷa	⁺tisilad	⁺tsahkar	⁺tsɨwd	⁺dimbaku
PMada	⁺mikʷaw	⁺łasila	⁺łalahkar	⁺abulumbaw	⁺dzumikʷ
PMofuᴳ	*mikʷaw	*tasila	⁺daŋgafad	⁺tsufwad	*kiraw
PMbukoᴳ	*marka	⁺tsɨwɓay	⁺nifudawfudaw	⁺disuwdaw	⁺kuraw
PHurzo	⁺marka	⁺tsɨwɓay	⁺dzi[ms]iray	⁺tahkay	⁺dzim
PMarwaᴳ	*markɨd	⁺tarna	⁺daŋgafad	⁺gwaltir	*kirɨʷ
PUldeme	⁺mikʷaw	⁺sisila	⁺fadfad	⁺ʔalɓit	⁺kʷilaw

Baka numerals by Roger Blench (2009b).

Baldamu numerals by Seignobos & Tourneux (1984: 27).

Cuvok numerals by Ndokobaï (2008) + (2009) https://mpi-lingweb.shh.mpg.de/numeral/Cuvok.htm

Dugwor numerals by Hubert Nkoumou (2007) https://mpi-lingweb.shh.mpg.de/numeral/Dugwor.htm

Giziga₁ – North numerals by Henry Tourneux (1995) https://mpi-lingweb.shh.mpg.de/numeral/Gizga-North.htm

Giziga₂ – South numerals by Erin Shay (2007) https://mpi-lingweb.shh.mpg.de/numeral/Gizga-South.htm

Giziga₃ numerals by Johannes Lukas (1970: 39).

Giziga₄ numerals by Kurt Strümpell (1923: 120-23).

Hurzo₁ = Gwendele numerals by Véronique de Colombel (letter from the Oct 20, 1987).

Hurzo₂ numerals by Jean Mouchet (1953: 194-95).

Mada₁ numerals by Véronique de Colombel (letter from the Oct 20, 1987).

Mada₂ numerals by Huber Nkoumou (2007) https://mpi-lingweb.shh.mpg.de/numeral/Mada.htm

Mada₃ numerals by Jean Mouchet (1953: 194-95).

Mafa₁ numerals by Henry Tourneux (1995) https://mpi-lingweb.shh.mpg.de/numeral/Mafa.htm
(cf. Barreteau & Le Bléis 1990: 52)

Mafa₂ = Matakam numerals by Charles Kraft (1981, III: 145).

Magumaz numerals by Véronique de Colombel (letter from the Oct 20, 1987).

Matakam₁ numerals by Kurt Strümpell (1923: 120-23).

Matakam₂ numerals by Kurt Strümpell (1923: 120-23).

Matakam₃ numerals by Jean Mouchet (1953: 194-95).

Matal numerals by Arjan Branger via Ginger Boyd (2013) https://mpi-lingweb.shh.mpg.de/numeral/Matal.htm

Mbuko₁ numerals by Richard Gravina https://mpi-lingweb.shh.mpg.de/numeral/Mbuko.htm

Mbuko₂ numerals by Roger Blench (2009b).

Mbuko₃ numerals by Jean Mouchet (1953: 194-95).

Mefele numerals by Ndokobaï (2008) https://mpi-lingweb.shh.mpg.de/numeral/Mefele.htm

Merey numerals by Henry Tourneux (1995) https://mpi-lingweb.shh.mpg.de/numeral/Merey.htm

Mikiri numerals by Roger Blench (2009b).

Mofu₁ numerals by Jean Mouchet (1953: 194-95).

Mofu₂ numerals by Véronique de Colombel (letter from the Oct 20, 1987).

Mofu-Gudur numerals by Kenneth R. Hollingsworth https://mpi-lingweb.shh.mpg.de/numeral/Mofu-Gudur.htm

Mofu – North numerals by Hubert Nkoumou (2007) https://mpi-lingweb.shh.mpg.de/numeral/Mofu-North.htm

Moloko numerals by Diane Friesen (2007, 2012) https://mpi-lingweb.shh.mpg.de/numeral/Moloko.htm

Mufo numerals by Kurt Strümpell (1923: 120-23).

Mufu numerals by Kurt Strümpell (1923: 120-23).

Muktele[1] numerals by Véronique de Colombel (letter from the Oct 20, 1987).

Muktele[2] = Balda numerals by Kurt Strümpell (1923: 120-23).

Muturua numerals by Kurt Strümpell (1910: 456-58).

Muyang[1] numerals by Véronique de Colombel (letter from the Oct 20, 1987).

Muyang[2] numerals by Tony Smith https://mpi-lingweb.shh.mpg.de/numeral/Muyang.htm

Uldeme[1] = Wuzlam numerals by Willie Kinnaird https://mpi-lingweb.shh.mpg.de/numeral/Ouldeme.htm

Uldeme[2] numerals by Veronique de Colombel (1997: 47).

Uzlam numerals by Jean Mouchet (1953: 194-95).

Vame numerals by Willie Kinnaird https://mpi-lingweb.shh.mpg.de/numeral/Vame.htm

Zelgwa[1] = Zulgo numerals by Véronique de Colombel (letter from the Oct 20, 1987).

Zelgwa[2] = Zulgo-Gemzek numerals by Beat Haller (1994) https://mpi-lingweb.shh.mpg.de/numeral/Gemzek.htm

Zelgwa[3] numerals by Jean Mouchet (1953: 194-95).

2.6. Sukur

Language	1	2	3	4	5
Sukur[1]	*kàlí*	*bák*	*mãken*	*fwáɗ*	*ɮám*
Sukur[2]	*kɪli*	*bak*	*makɪn*	*fwɔt*	*łam*
Proto-Sukur	*+kili*	**bak*	*+makin*	**fwaɗ*	**ɮam*

Language	6	7	8	9	10
Sukur[1]	*mókwà*	*mádáf*	*tòkòz*	*míçí / míłí*	*ʔwàn*
Sukur[2]	*mukwa*	*madaf*	*tɪgɪz*	*mɪxi (= mikhi)*	*wɔŋ*
Proto-Sukur	**mɨkʷa*	**maɗaf*	**tiɣis*	*+miłi*	**waŋ*

Sukur[1] numerals by Herrmann Jungraithmayr https://mpi-lingweb.shh.mpg.de/numeral/Sukur.htm

Sukur[2] numerals by Meek (1931: 318).

2.7. Daba

Language	1	2	3	4	5
Buwal	*téŋgʷōlèŋ*	*gɓák*	*māxkát‾*	*ǹfát‾*	*dzābán*
Gawar[1]	*taguleng*	*bak*	*mãchkát*	*fãt*	*tschĕbéng*
Mbedam	*ntaɗ*	*bak*	*maxkaɗ*	*mfaɗ*	*dʒəban*
Hina[1]	*ǹtá*	*suloɗ*	*mahkaɗ*	*mfáɗ*	*dzəbuŋ*
Hina[2]	*ta*	*solő*	*mākǎ́*	*fa*	*sēbő́n*
Hina[3]	*nta*	*sollo*	*mǎkǎ́*	*fa*	*sĕbɨ́n*
Gawar[2]	*takan*	*səray*	*makaɗ*	*faɗ*	*jeɓin*
Daba[1]	*tòkòn*	*sìrəy*	*màkat*	*fwot*	*jìɓun*

Daba₂	*takán*	*sorā*	*mākǎ́*	*fŏ́(d)*	*dschŏ́ben*
Daba₃	*tākǎ́n*	*sěrai*	*mākǎt*	*fŏ́(d)*	*jĭbĭ́n*
Musugeu	*takan*	*sray*	*makat*	*fwǫd*	*jə̆buŋ*
*	⁺*takan* ⁺*ntaɗ*	**ɓik* **sɨraj*	**mahkaɗ*	**wɨfaɗ =* ⁺*mɨfwaɗ*	**dzaɓɨn*

Language	6	7	8	9	10
Buwal	*ŋ̀ʷkʷáx*	*ŋ̀ʃílέ́t̚*	*dzāmāxkāt̚*	*dzáfát̚*	*wám*
Gawar₁	*nkŏ(ch)*	*schĭlέ́*	*jămăchkǎ́t*	*jǎfăt*	*wam*
Mbedam	*ŋkwax*	*diʃliɗ*	*dʒamaxkaɗ*	*tsafaɗ*	*wam*
Hina₁	*ǹkú*	*dìsùlùɗ*	*fádfáɗ*	*varkanta*	*gə̀b*
Hina₂	*nku*	*dsĭlέ́(t)*	*fādfād*	*berkantǎ́*	*gē*
Hina₃	*nkō*	*tsĭlé / dsilé*	*fadefát*	*bărkăntǎ́*	*gĕ(b)*
Gawar₂	*koh*	*cesireɗ*	*cəfadcəfad*	*dərfatakan*	*guɓ təɓa təɓa*
Daba₁	*kwox*	*cìsìre*	*cifətcifət*	*dìrfə̀tə̀kən*	*gup*
Daba₂	*kū(έ́)*	*dschoe serǎ́*	*tschötsch fǎ́*	*gērf gātākǎ́n*	*gō⁽ⁿ⁾*
Daba₃	*kō*	*sērét̚*	*sfatsfat*	*dērfǎtăkǎ́n*	*gŭ(b)*
Musugeu	*kuh*	*čisrey*	*čʊfatčʊfat*	*drefkataken*	*gup*
*	**kʷah =* ⁺*ŋkʷah*	**tsasaraɗ =* ⁺*tsa-sɨraɗ*	⁺*dzama-hkaɗ* ⁺*faɗfaɗ* ⁺*tsɨfaɗtsɨfaɗ*	⁺*dza[ma?]-faɗ* ⁺*dɨrfa-takan*	⁺*wam* ⁺*guɓ*

Buwal numerals by Michael & Melanie Viljoen (2007) https://mpi-lingweb.shh.mpg.de/numeral/Buwal.htm

Daba₁ numerals by Charles Kraft (1981, III: 156).

Daba₂ = Musugeu numerals by Hauptmann Strümpell (1910: 456-58).

Daba₃ numerals by Kurt Strümpell (1923: 120-23).

Gawar₁ numerals by Kurt Strümpell (1923: 120-23).

Gawar₂ numerals by Matthieu Dzia and Robert Hedinger (1994) https://mpi-lingweb.shh.mpg.de/numeral/Daba.htm

Hina₁ = Mina numerals by Hubert Nkoumou (2007) https://mpi-lingweb.shh.mpg.de/numeral/Mina.htm

Hina₂ numerals by Kurt Strümpell (1910: 456-58).

Hina₃ numerals by Kurt Strümpell (1923: 120-23).

Mbedam numerals by Michael & Melanie Viljoen (2007) https://mpi-lingweb.shh.mpg.de/numeral/Mbedam.htm

Musugeu numerals by Jean Mouchet (1950: 58-59).

2.8. Bata

Language	1	2	3	4	5
Mubi	*rŭngoǎ́*	*braie*	*mǎka*	*nfǎdă*	*tŭfna*
Gude₁	*tèen / rûŋ*	*bə̀rá↑y*	*màkk*	*ǹfwáɗ*	*tə̆f*
Gude₂	*tân*	*bray*	*mak*	*↑ùnfwəɗ*	*tuʃ*

	1	2	3	4	5
Bachama	*hiɗò*	*k͡pe*	*ḿwɔ̀kun*	*fwət*	*tuf*
Mwulyen	*hídò / tên*	*búk / bíkǯ*	*mwàkín / maxkə̀n*	*fwad / fwát'*	*túhf / túf*
Bata₁	*hidő*	*bae*	*mōākén*	*fōa̋(t)*	*tūf*
Bata₂	*hído̧*	*kpwě*	*mwakən*	*fwo̧t*	*tuf*
Bata-Garua	*jido*	*te*	*moakén*	*fōa̋t*	*tūf*
Bata-Demsa	*hído*	*pě*	*moaké̋n*	*fōa̋t*	*tūf*
Koboči	*hidi*	*baek*	*maukén*	*fōa̋t*	*tŭf*
Wadi	*hĭdŏ*	*bĕg*	*moakín*	*fōa̋t*	*tŭf*
Holma	*hido*	*beg*	*mōāke̋n*	*fōa̋t*	*tuf*
Nzanyi₁	*hɪdè*	*buk*	*mìdi̧fəl*	*fwət*	*tuf*
Nzanyi₂	*hĭdō*	*beg*	*mĕnfe̋n*	*fōa̋t*	*tuf*
Nzanyi₃	*hído̧*	*gbwək*	*mwő̧kən*	*fwat*	*tuf*
Gudu	*jǎŋ*	*bæ̀k*	*mā:kə́n*	*fwád*	*tùf*
*	⁺*tan* ⁺*hiɗaw*	**ɓijak* = ⁺*ɓiryak*	**mahikin*	**fʷaɗ*	**tifʷ*

Language	6	7	8	9	10
Mubi	*kőǎ*	*ma̋dĕfǎ*	*tĕgessa*	*linga*	*pǎ̋ǎ*
Gude₁	*kùwà*	*mɔ̀dəf*	*tɔ̀ɣɔ́s*	*ìllíŋ*	*puʔ*
Gude₂	*kuwà*	*midîf(a)*	*tìɣis(a)*	*ʔìlliŋ(à), ìlɗiŋ*	*pu(wà)*
Bachama	*tukwɔ̀ltaka* 5 + 1	*tukòluk͡pe* 5 + 2	*fwɔ̂fwət* 4 + 4	*dɔ̀mbidò*	*bɛ̆w*
Mwulyen	*túkwàldèáká / bɜ̀rfíŋ*	*túkwàlóʔpé / tùʁús*	*fwáfwàd* 4 + 4 / *mìʒíŋ*	*táàmbídò / pó?*	*bù*
Bata₁	*tūgoldaka*	*tugolpé*	*foafő̧t*	*tǎmbīdō*	*bō(ǎ)*
Bata₂	*to̧ko̧l dǎ̧ka*	*to̧ko̧l okpwe*	*fwatfwat*	*tambido*	*bōu*
Bata-Garua	*tokeldākă*	*tŏkŏlŏpe̋*	*fāfa̋̋t*	*tāmbĭddŏ*	*bo*
Bata-Demsa	*tūkŏldǎ̧kă*	*tu kŏlŏpe̋*	*fōa̋fōa̋t*	*tāmbĭdō*	*bū*
Koboči	*koach*	*mŭska̋̋tŭ*	*foafoade*	*tămbĭdē*	*pŭ*
Wadi	*kōa̋*	*miska̋ta*	*foafat*	*tămbillo*	*pū*
Holma	*kōa̋*	*măskáta*	*fōfoa̋ddĕ*	*tawidde*	*pu*
Nzanyi₁	*kwɔx*	*mìskatɔ̀*	*fwəfwadè*	*təmɓedè*	*pu*
Nzanyi₂	*kōa̋*	*mĕska̋tā*	*fofoade*	*tanbidde*	*pu*
Nzanyi₃	*kwah̠*	*miska̧ta*	*fo̧fwa̧dȩ*	*tambido*	*pu*
Gudu	*kwă*	*mīskàtā*	*fɔ̄rfwād*	*žĭétə́pə̀n*	*pú*
PBata	**kiwa* ⁺*tukʷil-taka*	**midîfiŋ* ⁺*tukʷil-ɓi[ry]ak* ⁺*miskata*	**tiɣis* ⁺*fʷadfʷaɗ*	⁺*ʔilȥiŋ* ⁺*tambiɗaw*	⁺*[g]paw*

Bachama by Russell G. Schuh (1989) https://mpi-lingweb.shh.mpg.de/numeral/Bacama.htm
Bata₁ numerals by Kurt Strümpell (1910: 456-58).
Bata₂ numerals by Jean Mouchet (1950: 58-59).
Bata-Dempsa numerals by Kurt Strümpell (1923: 120-23).
Bata-Garua numerals by Kurt Strümpell (1923: 120-23).
Gude₁ numerals by Russell G. Schuh (1989 https://mpi-lingweb.shh.mpg.de/numeral/Gude.htm
Gude₂ numerals by Charles Kraft (1981, III: 36),
Gudu numerals by Charles Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/Gudu.htm

Holma numerals by Kurt Strümpell (1923: 120-23).

Koboči numerals by Kurt Strümpell (1923: 120-23).

Mubi numerals by Kurt Strümpell (1923: 120-23) (a variety of Gude; ≠ Mubi of East Chadic origin)

Mwulyen = Jimi = Zumu numerals by Charles H. Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/Jimi.htm

Nzanyi[1] numerals by Charles Kraft (2007) https://mpi-lingweb.shh.mpg.de/numeral/Nzanyi.htm

Nzanyi[2] numerals by Kurt Strümpell (1923: 120-23).

Nzanyi[3] numerals by Jean Mouchet (1950: 58-59).

Wadi numerals by Kurt Strümpell (1923: 120-23).

2.9. Gidar – Musgu

Language	1	2	3	4	5
Gidar[1]	*tákà*	*súlà*	*hókù*	*pódò*	*łé*
Gidar[2]	*tĕtăkă*	*sūlă*	*hŏkū*	*pŏdŏ*	*ksē*
Mbara	*kítáy, dów*	*mòk*	*ùhú*	*púdú*	*iłím*
Munjuk	*kətay, daw*	*sulu*	*hu*	*pudu*	*slim*
Musgu[1]	*kítáy, dáw*	*súlú*	*hú*	*púdú*	*łím*
Musgu[2]	*katē*	*tšílo*	*hū*	*pŏdu*	*χim*
Musgu[3]	*daya*	*tsolő*	*u*	*poru*	*šem*
Musgu[4]	*kedái, ketái*	*silú*	*ħu*	*pudű*	*χ̃im, šĩm*
Musgu[5]	*ketē*	*tšựlo*	*hū*	*puddu*	*tim*
Musgu[6]	*dai*	*solú*	*hu*	*podú*	*tim*
Musgu[7]	*dai // degá*	*χillu // hudyú*	*hūū / hūuŋ*	*pudu // fūdi*	*sim // hrāo*
Muskum[1]	*degá*	*hujú*	*hūang*	*fūdi*	*hrāi*
Muskum[2]	*dỉgà*	*wùzìk*	*hú:wúŋ*	*fú:dì*	*hìrrày*
PGidar PMusgu PMuskum	+*takay*/+*katay* +*daway* +*dỉga*	*sỉwla *silỉ +*hilyu*?	+*haku*[n?] +*ʔuhu*[n?] +*huwaŋ* < +*hugaŋ*?	**pad̂a*[w] *pidỉ[w] +*fudỉ*	*łaʔ *ʔiłím +hiray

Language	6	7	8	9	10
Gidar[1]	*łré*	*bùhúl*	*dòdòpórò*	*váyták*	*kláù*
Gidar[2]	*sĕrrő*	*būl (bwul)*	*dĕdĕpŏdŏ*	*waităk*	*kēlaú*
Mbara	*łírá*	*mìgzàk / mùgizàk*	*mìsílày / mùsílày*	*wá:ŋá*	*dò:gò / dòk*
Munjuk	*slaara*	*mə(gə)zak*	*mit(i)wis*	*təkla / zleeŋe*	*doogo*
Musgu[1]	*là:rà*	*mìgzàk / mùgzàk*	*mìtwìs / mìtìs*	*tíklá*	*dò:gò*
Musgu[2]	*χ̃āra*	*mugzāk*	*mitíš*	*ỹēŋgar*	*dógō*
Musgu[3]	*saãra*	*nogosap*	*meteš*	*takala*	*tor*
Musgu[4]	*χ̃āra*	*múkezak*	*métūiš*	*dékela*	*dőgo*
Musgu[5]	*tará*	*mugzak*	*mutíš*	*tikela*	*dōgo*
Musgu[6]	*tara*	*mgzek*	*mtiš*	*ldenge*	*dóko, dogo*

Musgu[7]	sāra//hrdíga	mugezák//māde	mutuīš // hasgaŋ	tíkilla // hrfidi	ndōgo // gum
Muskum[1]	hrdíga	mādi	hasgang	hrfidi	gúru
Muskum[2]	hìrdî:gà	mà:dî	hà:skàŋ	hírífídî	kúrú
PGidar PMusgu PMuskum	*łira *łira +hiray-dîga	+busɨwla +mukɨzak cf. Mbara mòk 2 Muskum[2] wùzɨk 2 +maadî	+dadapaɗaʷ +muti[ɣ]iš +musɨlay +haas-gaŋ cf. +hugaŋ 3	+vaytak +tɨkila +ǯiyaŋay +hiray-fudî	+kɨlaw +dogo +kuru

Gidar[1] numerals by Zygmunt Frajzyngier (1996) https://mpi-lingweb.shh.mpg.de/numeral/Gidar.htm
Gidar[2] numerals by Hauptmann Strümpell (1910, 456-58).
Mbara numerals by Henry Tourneux (1988) https://mpi-lingweb.shh.mpg.de/numeral/Mbara.htm
Munjuk numerals by Henry Tourneux (1982).
Musgu[1] = Mulwi = Vlum = Mogroum numerals by Henry Tourneux (1977: 20-30; 1988)
https://mpi-lingweb.shh.mpg.de/numeral/Musgu.htm
Musgu[2] numerals by Heinrich Barth apud Johannes Lukas (1941).
Musgu[3] ('Masa') numerals by Decorse apud Johannes Lukas (1941); cf. Gaudefroy-Demombynes
(1907: 249).
Musgu[4] numerals by Krause apud Johannes Lukas (1941).
Musgu[5] numerals by Overweg apud Johannes Lukas (1941).
Musgu[6] numerals by Rohlfs apud Johannes Lukas (1941).
Musgu[7] numerals by Röder apud Johannes Lukas (1941).
Muskum[1] = Musgum City numerals by Johannes Lukas (1937: 143).
Muskum[2] numerals by Tourneux (1977: 20-30).

2.10. Kotoko

Language	1	2	3	4	5
Buduma[1]	gə̀tté	kí	gàkə́nnə́	hígáy	híŋɪ̀
Buduma[2]	kẹ́ttè	kì̃	kākénnẹ	hǎ́γaɪ	sinjí
Buduma[3]	kéta	kĩ́hi	kǎ̃kènè	hérai, sérai	hĩ́ndʃi
Buduma[4]	kittā	kihī	kakínne	segay	sínji
Kuri	kete	kin	kakenne	kaï	šinǰi
Ngala[1]	takshán	kĕshē	kînge	kādi	khathaezi
Ngala[2]	tiksang	kisang	kinga	kadi	kishenshi
Małam	ǹté	gàsíy	gákə̀rò	gàdè	ɬ'éŋsì
Afade[2]	ǹté	gàsíy	gákùrò	gàdè	ɬ'èntsì
Afade[3]	pal	gasi	kákuro	kadẹ	şẹnsi
Afade[4]	te	anszih	ankró	gandéh	täntih
Afade[5]	dīpan	wagāsi	gákŭru	wagāde	wasēntshi
Sahu	ǹtè	gáŋsìy	gá:kə̀rà	gàndě	c'èŋsí
Mpade = Makeri[1]	pál	gāsì	gòkúrò	gādè	ʄénsī
Mpade = Makeri[2]	ǹté	gàsìy	gòkúrò	gàdè	šéŋsì
Makeri[3]	pal	gasi	gọkurọ	gandẹ	şyẹnsi
Shoe	nte	ansi	ankro	gade	sesi
Gulfei[1] = Malgbe	ǹté	'èyá	'àkrà	gàndě	'éŋsì / ɬéŋsì

Gulfei₂	enté	ēá	ankrá	gandé	aínse
Gulfei₃	nte	er̠	akra	ngandé	ensi
Gulfei₄	èn̠té	èyá	àkrà	gān̠de	ēnsi
Gulfei₅	pal	gęn̠	gak(ə)r	gandę	şensi
Kuseri₁ = Mser	s'ə́ɣə́dìy	kícó	kákə̀r	kádę̆	šèší
Kuseri₂	tsegiden	kičǫ	kagər	kadę	sęsi
Kuseri₃	segedi	kítio	kák̀r	kǻde	sęsi
Huluf	s'ə́ɣə́dìy	kísów	káxkə̀r	kádę̆	šèší
Kala-Kafra	s'ə́ɣə́dìy	kísó	ká:k̀r	kádę̆	sèsí
Afade₁	sə́rə̀jā	sdā	gàrkə̀	gàdē	ʃìʃí
Lagwan₁	sə́ɣdia, tkú	χsdǻ	gə̆χkər	gə̆de	ʃẽʃí
Lagwan₂	sə́ɣə́dìyà	χsdę̆	gáxhə̀r	gádę̆	šèší
Lagwan₃	sə́ɣdia, tkú, bal ᴺ	χsdǻ	gáχəkər	gáde	šēší
Lagwan₄	tekū; serédiā	ksdē	gáχkir	gāde	sēsi
Zina = Jəyna	cíyá	cúw	hòkwá	fowdìy	hə̀rè
Mazera	ńcíyàn	ńcò	hùnò	fùdę̆	'ìhrę́
Ngodeni	ǹs'á	ǹcó	'àxkwán	fò:dí	hxrè
Mo'e	ǹs'á	ǹcó	'àxkwán	fò:dí	hrè
PKotoko PNKotoko PCKotoko PIsland Kotoko PSKotoko	⁺tik / ⁺tay / ⁺pal ⁺tiku / ⁺tsi-ɣidiya / ⁺bal ⁺kittay ⁺tsiya	*kasi *gasi *xisa *kihi *tsu	⁺gakɨraw ⁺kaxikir ⁺kakinnay ⁺xɨkʷan	⁺ka-fday ⁺gadę̆ = ⁺gad̆ay ⁺gad̆ay ⁺hɨgay *fodî	*ɬensi = ⁺ɬentsi *ɬ'entsi *ʃeʃî *hiᵞgi = ⁺hinǰi ⁺xihray

Language	6	7	8	9	10
Buduma₁	*hə̀ràkkə̣́*	*tùlwár*	*wósə̣́kə̣́*	*hílígár*	hákkán
Buduma₂	*hāráke*	*túlor*	*fɔkú*	*híligar*	hākán
Buduma₃	*hárasgè, sáraskè*	*tulṓr*	*hūásgè*	*héligār*	hákan
Buduma₄	*sarásko*	*ty̆lór*	*sōsku*	*selĭyáŋ*	sĕkan
Kuri	*saraske*	*tolor*	*soske*	*(fi)liger*	sekkan
Ngala₁	*rasko*	*tyllur*	kēdi gădī	nadăkhsáh	hăkhkán
Ngala₂	kingi ti kisang	tiksang ala gada	kadi kadi	ina tiksang	hakang
Małam	*frákə̀rò*	*dúlò*	gàdē̆-gàdē̆	nòlké	kàŋ
Afade₂	*frâkúró*	*dúlò*	gàdē̆-gàdē̆	nótè	kàŋ
Afade₃	*frákurǫ*	*dulǫ*	gadę̆gadę	nǫtę	kãg
Afade₄	*frâkro*	*dûllo*	kadégadéh	nottéh	dékang
Afade₅	*wafrákĕrō*	*adūro*	wagadégădē	wánŏte	wăkán
Sahu	*fəràkə̀rà*	*dùlù*	frègàndè	dáyy'à	kàŋ
Mpade = Makeri₁	*ʃèskótē*	*túlùr*	jìlìgàdè	jìàtálà	kán
Mpade = Makeri₂	*šéskótè*	*túlûr*	jìlì-/jìdì-gàdè	jàtállà	kàŋ
Makeri₃	*şyęskǫte*	*tulu*	gădę̆gadę	tyatalla	kãg

Shoe	seskote	tulur	gedégadé	deatala	kang
Gulfei = Malgbe	frékrà	tùllúr	frègàndè	díyáyá	kàŋ
Gulfei₂	frékra	tulúr	fregandé	diańa	kān
Gulfei₃	frekra	tulur	fregande	ndiaga	kang
Gulfei₄	fẹ́rékrà	tùlúr	fẹregānde	dīań́a	kaŋ
Gulfei₅	frẹk(ə)r	tullur	fregundẹ	tyaya	kãg
Kuseri = Mser	vràkàkə̀r	kàtól	vrèkádẽ	nòrké	kàn
Kuseri₂	vrẹkadẹ	katəl	vrẹkadẹ	nọrkẹ	kan
Kuseri₃	wukákr	kə́tr	vrēkádē	núrgé	kān
Huluf	vràkáxkə̀r	kàtól	vrèkádẽ	nòtké	xkàn
Kala-Kafra	vràkákə̀r	kàtól	vrèkádẽ	nòtké	kàŋ
Afade₁	və̀nārkə̄	kàtùl	vìyādẽ	dìʃẽ̄	χkàn
Lagwan₁	vɛnăχ(ə)kər	kátul	vɛnáde	diʔiʃén < Ar tisʕa 9?	χkan
Lagwan₂	və̀náxkə́r	kátúl	vìṇádẽ	dìsìdĕn	xkàn
Lagwan₃	vɛnáχəkər	kə́tul	veńáde	di'išén	χkan
Lagwan₄	venáχkir	kātul	venyāde	dísχiēn	χkán
Zina = Jəyna	hàrkə́ncíyá	ɣwàdàl	màɣ̀ràdã́	cíyágàkə̀	lə̀ɓàkù
Mazera	ʼìbúwè	mùsxwámè	fúdàdè	fə̀rcádã̀	ʼìsàʼəm
Ngodeni	hxrè gə̀m sʼá	hxrè gə̀m có	jàŋ fòdíy	màgànà húné sʼá	lə̀ɓàkù
Moʼe	hrè gə̀m sʼá	hrè gə̀m có	jăŋ fódíy	màgànà húní sʼá	lə̀ɓàkù
PKotoko PNKotoko PCKotoko PIsland Kotoko PSKotoko	*vinahkir *vinahkir & ⁺ɬʼentsi-ku-tay *vinahkir < Kan. araskà ⁺xihray-kim-tsiya 5+1	< Kan. túlùr < ⁺ka- + túlùr < Kan. túlùr ⁺xihray-kim-tsu 5+2	⁺gadaygaday ⁺vinagaday < Kan. wuskú ⁺dzaŋ fodî & ⁺foda-foday	⁺inu tay finger 1 ⁺taya inaw finger 1 ⁺inu tiku finger 1 < Kan. ləgár ⁺tsiya ... (minus 1)	*kan *hikan *hikan ⁺liɓaku

Afade₁ numerals by Dakinodji Ngarmadjibé (1993 https://mpi-lingweb.shh.mpg.de/numeral/Afade.htm
Afade₂ numerals by Henry Tourneux (2005).
Afade₃ numerals by Jean-Paul Lebeuf (1942: 172).
Afade₄ numerals by Ulrich J. Seetzen (beginning of the 19th cent.) apud Heinz Sölken (1967: 157, 173-178).
Afade₅ numerals by Heinrich Barth (1912; see Heinz Sölken (1967, 165-69).
Buduma₁ numerals by E. Ari Awagana (2007) https://mpi-lingweb.shh.mpg.de/numeral/Buduma.htm
Buduma₂ numerals by Johannes Lukas (1939: 65).
Buduma₃ numerals by Sigismund W. Koelle (1854) apud Heinz Sölken (1967: 165-69).
Buduma₄ numerals by Heinrich Barth (1912; see Heinz Sölken (1967, 165-69).
Gulfei₁ = Malgbe numerals by Henry Tourneux (2005).
Gulfei₂ numerals by Adolf Friedrich apud Johannes Lukas (1937: 146).
Gulfei₃ numerals by Otto Röder apud Johannes Lukas (1937: 146).
Gulfei₄ numerals by Johannes Lukas (1937: 146).
Gulfei₅ numerals by Jean-Paul Lebeuf (1942: 172).
Huluf numerals by Henry Tourneux (2005).

Kala-Kafra numerals by Henry Tourneux (2005).

Kuri numerals by Decorse apud Maurice Gaudefroy-Demombynes (2007: 276).

Kuseri[1] = Mser numerals by Henry Tourneux (2005).

Kuseri[2] numerals by Jean-Paul Lebeuf (1942: 172).

Kuseri[3] numerals by Johannes Lukas (1937: 144).

Lagwan[1] numerals by Sean Allison (2007) https://mpi-lingweb.shh.mpg.de/numeral/Lagwan.htm

Lagwan[2] numerals by Henry Tourneux (2005).

Lagwan[3] numerals by Johannes Lukas (1936: 53).

Lagwan[4] numerals by Heinrich Barth (1862: 9, 11).

Makeri[3] numerals by Jean-Paul Lebeuf (1942: 172).

Małam numerals by Henry Tourneux (2005).

Mazera numerals by Henry Tourneux (2005).

Mo'e numerals by Henry Tourneux (2005).

Mpade = Makeri[1] numerals by Sean Allison (2007) https://mpi-lingweb.shh.mpg.de/numeral/Mpade.htm

Mpade = Makeri[2] numerals by Henry Tourneux (2005).

Ngala[1] numerals by Heinrich Barth (1912); see Heinz Sölken (1967: 165-69).

Ngala[2] numerals by Frederick W.H. Migeod (1922); see Heinz Sölken (1967: 165-69).

Ngodeni numerals by Henry Tourneux (2005).

Sahu numerals by Henry Tourneux (2005).

Shoe numerals by Johannes Lukas (1937: 154).

Zina = Jəyna numerals by Henry Tourneux (2005).

2.11. Masa

Language	1	2	3	4	5
Herdé	*ɗãw*	*hʷóèɓ*	*hĩ́ndʒì?*	*fídî?*	*vàl*
Lame[1]	*ɗáwà*	*hʷóɓò*	*hínčì?i*	*fùdî?i*	*vɪ́l*
Lame[2]	*ɗăó*	*hōɓō*	*híndzì?ì*	*fídî?*	*vàłà*
Lame[3]	*do*	*wobu*	*hinzi*	*fodi*	*war*
Mesme	*ɗãw*	*hɔ̀ɓ*	*hīndì*	*fídí*	*vàtl*
Peve[1]	*ɗaw*	*hoɓ*	*hínji?*	*fə́dî?*	*vál*
Peve[2]	*ɗao*	*hwōɓ*	*hínjī*	*fúdî*	*vātl*
Dari	*dongbǎ*	*hombǎ*	*hinji*	*fudi*	*was*
Musey[1]	*dèw*	*ɓà*	*híndí*	*fídí*	*fàl*
Musey[2]	*dèw*	*mbà*	*hìndì*	*fídí*	*vàl*
Musey[3]	*dew*	*mba*	*hindi*	*fidi*	*fahl*
Marba[1]	*tú*	*mbà*	*híndî*	*fídî*	*vál*
Marba[2]	*tù*	*mbà*	*hìndí*	*fídí*	*vàl*
Masana	*kèp, tù, tùm*	*mà?*	*hìdí*	*fidî*	*vàl*
Masa	*tum*	*ma?*	*hidi*	*fidî*	*faθ*
Gumay	*tum*	*ma?*	*hidi*	*fidî*	*fał*
Bana	*tum*	*ma*	*hidi*	*fidi*	*faš̄ch*
Banana[1]	*tìyəw*	*mbà*	*yìntì(di)*	*fidi*	*fàl*
Banana[2]	*tṹ*	*mbà*	*híndì*	*fídí*	*fàš̃*

Kulung	*turúm*	*mbá*	*hindi*	*fidi*	*waš*
PMasa (Shryock) PSMasa PNMasa	⁺*ɗaw* ⁺*tu*	⁺*humɓaw* ⁺*mɓaw*	**hindi*	⁺*fuɗî* *fiɗî	**vał*

Language	6	7	8	9	10
Herdé	*kāŋgíʔ*	*sēdã̄*	*tʃɔhòʔ*	*tēfer̃dɛ̄w*	*gùɓ*
Lame₁	*kánki*	*šíːsà*	*čóhò*	*čəfʌdíʔò*	*gʷúɓú*
Lame₂	*káŋgīdĩ*	*sédã̄ʔã*	*tsóhō̃ʔō̃*	*téfērdàò*	*gùɓù*
Lame₃	*kanki*	*sieda*	*cióobo*	*cefodo*	*gu*
Mesme = Zime	*kāndī*	*sēdã̄*	*tʃɔhȭ*	*tērfĩdɛ̄w*	*gùɓ / gùp*
Peve₁	*kánkiʔ*	*syédaʔ*	*tsóhoʔ*	*tʃéfaɖew*	*guɓ*
Peve₂	*kánkí*	*sédā*	*tʃóhō*	*tʃĕfādēo*	*gwúɓ*
Dari	*kanki*	*schaeda*	*tschohó*	*téfĕdō*	*gu*
Musey₁	*kárgìyá*	*kídísìyá*	*kálvàndì*	*łèŋŋè*	*dògò*
Musey₂	*kàrgìjá*	*kìdìzìjá*	*kàlvàndì*	*ɮèŋè*	*dòk / dògò*
Musey₃	*karkiya*	*kidsiya*	*galavandi*	*hleŋe*	*doogo*
Marba₁	*kárgéyá*	*kídîzíyà*	*ʔàklávándí*	*łéŋá*	*dógò*
Marba₂	*kàragàyà*	*sìdìzìjá*	*klàvàndì*	*ɮèèŋà*	*dòk / dògò*
Masana	*kàrgìjà*	*sìdìjà*	*glàvándí*	*ɮèŋè*	*dòòk*
Masa	*karkia*	*sidia*	*glavandi*	*θeŋe*	*dogo / doogo*
Gumay	*karkiya*	*sidîya*	*glavandi*	*łɛːŋɛ*	*dɔːgɔ*
Bana	*kárkia*	*zídia*	*kelafánti*	*thēne*	*dógo*
Banana₁	*kàr̃gìya*	*kìdìsìya*	*kàlàvandè*	*łiyeŋ*	*tohogò*
Banana₂	*kárgáyá*	*kidisia*	*àkéḷáfándí*	*ʃ̃ẽnè*	*dogo*
Kulung	*kareia*	*kidesia*	*keláfti*	*sénga*	*dógo*
PMasa (Shryock) PSMasa PNMasa	**kargi* ⁺*kanki* ⁺*karki*	**siɗa* ⁺*siɗa* ⁺*sidîya*	⁺*tsuhawʔaw* ⁺*kalavandi*	⁺*tifardîʔaw* ⁺*ɮiyaŋa*	**gup* ⁺*dogo*

Bana numerals by Johannes Lukas (1937: 129).
Banana₁ numerals by Charles Kraft (1981, III: 178).
Banana₂ numerals by Johannes Lukas (1937: 134).
Dari = Weimba numerals by Kurt Strümpell (1910: 457-58).
Gumay numerals by Henry Tourneux (1977: 20-30).
Herdé numerals by Pierre Court (2008) https://lingweb.eva.mpg.de/channumerals/Herde.htm
Kulung numerals by Johannes Lukas (1937: 137).
Lame₁ numerals by Charles H. Kraft (1981, III: 197).
Lame₂ numerals by Michka Sachnine (1982).
Lame₃ numerals by Johannes Lukas (1937: 134).
Marba₁ numerals by Baktara Vangtou through Souleymane Kaddo (2014)
https://lingweb.eva.mpg.de/channumerals/Marba.htm
Marba₂ numerals by Antonino Melis https://lingweb.eva.mpg.de/channumerals/Marba.htm

Masa numerals by Charles Kraft (1981, III: 166).

Masana numerals by Antonio Melis https://lingweb.eva.mpg.de/channumerals/Masana.htm

Mesme = Zime numerals by Charles Kraft https://lingweb.eva.mpg.de/channumerals/Mesme.htm

Musey₁ numerals by Douna Wandi Bernard through Dr. Souleymane Kaddo https://lingweb.eva.mpg.de/channumerals/Musey.htm

Musey₂ numerals by Antonino Melis https://lingweb.eva.mpg.de/channumerals/Musey.htm

Musey₃ numerals by Charles Kraft (1981, III: 186).

Peve₁ numerals by Lazare Wambadan through Erin Shay https://lingweb.eva.mpg.de/channumerals/Peve.htm

Peve₂ numerals by Charles Kraft https://lingweb.eva.mpg.de/channumerals/Peve.htm

Summarisation of results

	1.1	1.2	1.3	1.4	1.5	1.6
Tera						
PTera	⁺*hiʔarda*					
PHwona		⁺*tital*				
Bura-Margi			⁺*taŋ*	⁺*tɨku*	⁺*duku*	⁺*dzaŋaw*
Bana			⁺*tanay*	⁺*tix*		
Lamang-Mandara						
PLamang		⁺*tala*		⁺*tɨkʷɨ*		
PMandara						
PPodokwo				⁺*kʷitɨrʔ*		
Mafa-Mada						
PMuktele				⁺*tɨkʷ-ɨla*		
PZelgwa						
PMafa						
PMefele						
PMada				⁺*fɨ-tɨk*		
PMofu				⁺*tɨkʷ*		
PMbuko				⁺*-tɨkʷ*		
PHurzo						
PMarwa						
PUldeme						
Sukur						
Daba				⁺*takan*		
Bata	⁺*hidaw*					Gudu *jə̆ŋ*
Gidar-Musgu						
PGidar				⁺*takay*		
PMusgu						
Muskum					⁺*dɨga*	

Kotoko						
PNKotoko				⁺tik		
PCKotoko				⁺tiku		
PIsland Kotoko						
PSKotoko						
Masa						
SMasa						
NMasa						
Notes				CCh ⁺tik-aw/-ay cf. ECu *tak-kaw / -ta(y) NOm: Kafa tokki / tok together with		

	1.7	1.8	1.9	1.10	1.11	1.12
Tera						
PTera						
PHwona						
Bura-Margi	⁺pału					
Bana	⁺paław	⁺lim				
Lamang-Mandara						
PLamang	⁺palaw		⁺din			
PMandara	⁺palaw					
PPodokwo						
Mafa-Mada						
PMuktele						
PZelgwa				⁺ʔilik		
PMafa						⁺sitaɗ
PMefele						
PMada						
PMofu	⁺pal					
PMbuko						
PHurzo						
PMarwa	⁺pal		⁺den		⁺ɓilay	
PUldeme					⁺ɓila	
Sukur				⁺kili		
Daba						
Bata						

Gidar-Musgu						
PGidar						
PMusgu						
Kotoko						
PNKotoko	⁺*pal*					
PCKotoko					⁺*bal*	
PIsland Kotoko						
PSKotoko						
Masa						
SMasa						
NMasa						
Notes	cf. CCu: Awngi *əmpál* 1; but cf. also Saharan: Kanuri *fal* 1			cf. Maban: Maba *illek* 'that one'	cf. CSud: Gbaya **ɓälä** 1	cf. ECh: Mokulu *sòo* 1 Om **ʔis-* 1 Sem **ʕaštay-* 1

	1.13.	1.14.	1.15.	1.16.	1.17.	1.18.
Tera						
PTera						
PHwona						
Bura-Margi						
Bana						
Lamang-Mandara						
PLamang						
PMandara						
PPodokwo						
Mafa-Mada						
PMuktele						
PZelgwa						
PMafa						
PMefele	⁺*(ʔa)mita*					
PMada						
PMofu						
PMbuko						
PHurzo						
PMarwa		⁺*sala*				
PUldeme		⁺*šeleŋ*				
Sukur						

Daba			+ntaɗ			
Bata						
Gidar-Musgu						
PGidar						
PMusgu				+ɗaway	+tay	
Kotoko						
PNKotoko					+tay	
PCKotoko						+tsi-ɣidiya
PIsland Kotoko					+kittay	
PSKotoko						+tsiya
Masa						
PSMasa				+ɗaw		
PNMasa					+tu < +tiyaw	
Notes	ECu *mat-/*mit- 1 Om: Gimira *maṭ- 1; cf. Kafa mittoo sole			cf. HsGw +ɗaya 1; CSud: Lutos ɗóí 1 Kaba Deme ɗíyá		

	2.1	2.2	2.3	2.4	2.5	3.1	4.1
Tera							+fʷaɗ
PTera	+rap					+kunu	
PHwona		*sɨrri				*mahkan	
Bura-Margi		+sɨɗayaw		+miłu		+maakiru	*fʷaɗu
Bana		+sul			+ɓɨwaku	*maxkin	+fwaɗɨ
Lamang-Mandara							
PLamang		+hisɨra				*hikina	*wifaɗ
PMandara				+mitsay	*bɨwa	+hikirɗa	*ʔufadɨ
PPodokwo		+ʔisɨra				+makɨra	+ʔufaɗa
Mafa-Mada							
PMuktele		+sɨla				+makɨr	+ʔufaɗ
PZelgwa		+sɨlaw				+makɨr	+ʔifaɗ
PMafa			+tsaw			*mahkar	*faɗ
PMefele			+(ts)atsaw			*mahkar	+fiwaɗ
PMada		+sɨla				*mahkar	+wifaɗ
PMofu		*sɨwla				*mahkɨr	*wifaɗ
PMbuko			*tsaw			*maakan	*fuɗaw
PHurzo			+tsiyaw			+máŋgan	+fuwɗaw
PMarwa			*tsɨw			*maakaŋ	*mufaɗ

PUldeme			+biri-tsiw		+biri-tsiw	+makar	+mifaɗ
Sukur					*bak	+makin	*fwaɗ
Daba		*siraj			*ɓik	*mahkaɗ	+mifwaɗ
Bata					+ɓiryak	*mahikin	*f^{w}aɗ
Gidar-Musgu							
PGidar		*siwla				+haku^n?	*pada^w
PMusgu		*sili				+ʔuhu^n?	*pidɨ^w
Muskum		+hilyu?				+hugaŋ?	+fudî
Kotoko							+ka-fday
PNKotoko		*gasi				+gakɨraw	+gaɗay
PCKotoko		*xisa				+kaxikir	+gaɗay
PIsland Kotoko		*kihi				+kakinnay	+higay
PSKotoko			*tsu			+xik^{w}an	*fodî
Masa						*hindi	
PSMasa					+humɓaw		+fudî
PNMasa					+mɓaw		*fidi
Notes	cf. BTg +rap NBc +rabu SBc +rwap ECh: Mawa *rap* 2; Jarawa *rwap* 2	CCh *siwra < PCh +ciraw 2; cf. SCu *tsada 2 ECu: Arb *sada* Yaaku *çɛ*	CCh *tsijiw; cf. Sah: Daza *čú:* Teda *cú:* Zagh *súyi* Bideyat *sui* Berti *su* 2	cf. ECh Sokoro *módû* 2 Ubi *mudû*	cf. Jrw Bantu +ɓari 2 NS: Maban +mbar 2	CCh *hikin PCh +kanu cf. PSOm *makkan 3?	CCh *wipad < PCh +fwad < +far(i)du; Beja *faɗig / farig* 4 ECu *ʔafur- / *ʔaffar- 4 POm *ʔawurd- 4 Egyptian *fd-w/-t* 4 < *fida3wat

	5.1	5.2	5.3	5.4	5.5	5.6	5.7
Tera							
PTera	*dîrman						
PHwona		+tif^w					
Bura-Margi		*tif^w					
Bana		+mits^{j}ifi					
Lamang-Mandara							
PLamang		*h^{w}itaf					
PMandara				+ʔiʐiɓa			
PPodokwo			+ʒama				
Mafa-Mada							

PMuktele				+ǰiw		
PZelgwa			+ʔiǰam			
PMafa			*ǰam			
PMefele			+ǰam			
PMada			+ǰam			
PMofu			*ǰim			
PMbuko		+ɗara(m)				
PHurzo		+ɗara				
PMarwa			*ǰinᵈdam			
PUldeme			+ǰam			
Sukur			*ǰam			
Daba				*dzaɓin		
Bata		*tifʷ				
Gidar-Musgu						
PGidar			*łaʔ			
PMusgu			*ʔiłim			
Muskum					+hiray	
Kotoko			*łentsi			
PNKotoko			*ł'entsi			
PCKotoko			*ʃeʃi			
PIsland Kotoko			+hinǰi			
PSKotoko					+xihray	
Masa						*vał
PSMasa						
PNMasa						
Notes		CCh *hʷitif; WCh *tafa palm of hand; Sokoro *tafa; CCu: Awngi *tafa hand	CCh *ǰiɗim < *ǰimi-ɗ & *łensi < +ʃemi(S); cf. Berb *sammūs 5		cf. CCh *dziviʸ arm; or Adm: Pana ndiɓi 5 Mambai bìzépè̜ 5	cf. CCh *hira arm, hand cf. WCh: Ron +haRa 5	WCh *bV-baĉu ECh *baẑyaw 5

	6.1 = 3+3	**6.2** = 2x3	**6.3** = (5)+1 or 5+(1)?	**7.1** = 3+4?	**7.2** = (5)+2?
Tera				⁺*mɨdu*	
PTera	⁺*ŋyoŋ*				
PHwona	⁺*miki*		Boka *tyɛ̀xxɛł*		
Bura-Margi	⁺*ŋkʷa*			**mɨdîfaw*	
Bana	⁺*ŋkʷaŋ*			**ᵐbirfiŋ*	
Lamang-Mandara					
PLamang	**mkɨwa*			⁺*rdfaŋ*	
PMandara	**mikɨwah*			**mɨdîfɨ*	⁺*vabɨwa*
PPodokwo	⁺*mikɨwa*			⁺*madifa*	
Mafa-Mada					
PMuktele	⁺*mukʷa*			⁺*mɨdîf*	
PZelgwa			⁺*ndilik*		⁺*tisila*
PMafa	**makʷa*				**tsarad*
PMefele	⁺*mikʷa*				⁺*tisiladʼ*
PMada	⁺*mikʷaw*				⁺*łasila*
PMofu	**mikʷaw*				**tasila*
PMbuko	**marka*				⁺*tsiwɓay*
PHurzo	⁺*marka*				⁺*tsiwɓay*
PMarwa	**markɨd*				
PUldeme	⁺*mikʷaw*				⁺*sisila*
Sukur	**mikʷa*			**madaf*	
Daba	⁺*ŋkʷah*				⁺*tsasirad*
Bata	**kiwa*		⁺*tukʷɨl-taka*	**mɨdîfiŋ*	⁺*tukʷil--ɓi[ry]ak*
Gidar-Musgu					
PGidar			**lira*		⁺*busɨwla*
PMusgu			**lira*		⁺*mukɨzak*
Muskum			⁺*hiray-dîga*	⁺*maadî*	
Kotoko					
PNKotoko		**vinahkɨr*	⁺*ł'entsi-ku-tay*		
PCKotoko		**vinahkɨr*			
PIsland Kotoko					
PSKotoko			⁺*xɨhray-kɨm--tsiya*		⁺*xɨhray-kɨm-tsu*
Masa					
PSMasa	⁺*kanki*				⁺*sida*
PNMasa	⁺*karki*				⁺*sidɨya*

| Notes | ⁺(mɨ-) kʷan-kʷan < ⁺(mɨ-) kanu-kanu | | | | ⁺mɨfɨdîkn < ⁺mɨ-fadî-kanu 4+3; cf. WCh: SBc: Dott wúsúr-maaká 7 = wópsi 4 + maaká 3 ECh: Ndam wosubo 7 = woro 4 + supu 3 | cf. CCh *tasɨrad |

	7.3	**7.4**	**8.1** = 4+4	**8.2** = ? 4	**8.3**	**8.4** = (10)-2	**8.5** = (5)+3
Tera							
PTera						⁺mɨyasɨri	
PHwona			⁺ fʷadfʷad				
Bura-Margi						*tsʲisɨw	
Bana					*tɨɣis		
Lamang-Mandara					*tɨɣis		
PLamang							
PMandara							
PPodokwo						⁺ɨzza < ⁺ɨsɨra dʒim 10-2?	
Mafa-Mada							
PMuktele					⁺mɨtigiš		
PZelgwa							⁺tsamakɨr
PMafa							⁺tsamakard
PMefele							⁺tsahkar
PMada							⁺łalahkar
PMofu				⁺daŋgafad			
PMbuko			⁺nɨfudaw-fudaw				
PHurzo						⁺dzi[ms]iray	
PMarwa	⁺tarna			⁺daŋgafad		Bald sàabùr	
PUldeme			⁺fadfad				
Sukur					*tɨɣis		

Daba		⁺fadʄad ⁺tsifadîtsi-fad				⁺dzamahkad
Bata	⁺miskata	⁺fʷadʄʷad		*tiɣis		
Gidar-Musgu						
PGidar				⁺muti[ɣ]iš		
PMusgu			⁺dadapadaʷ		⁺musilay	
Muskum						⁺haasgaŋ
Kotoko						
PNKotoko		⁺gaday-gaday				
PCKotoko			⁺vinagaday			
PIsland Kotoko						
PSKotoko		⁺fodafoday	⁺dzaŋ fodî			
Masa						
PSMasa					⁺tsuhawʔaw?	
PNMasa			⁺kalavandi?			
Notes	cf. Adm: Mambai *tàrnã̀gà* 7	cf. CSud: Kaba Na *mìtə́kə́jə́* 7		cf. WCh: Ngizim *dándà-fə́dú* 8	CCh *tiɣis* cf. HsGw *takwasi* < *biyat-akwan-si* 5 + 3	

	9.1 = (5)+4	9.2 = (10)-1	9.3	9.4	9.5	9.6
Tera						
PTera			⁺miɮam			
PHwona	⁺fʷad dîrman	⁺gwaŋ-ha-hiʔarda				
Bura-Margi			⁺miɮaw			
Bana			⁺miɮi			
Lamang-Mandara						
PLamang	⁺timbaday					
PMandara	⁺visulimbaday			⁺masil-manay		
PPodoko		⁺matirtsay				
Mafa-Mada						
PMuktele		⁺ladiga				
PZelgwa	⁺tsuwɨd					
PMafa	⁺tsuwad					
PMefele	⁺tsɨwd					

PMada						⁺ʔaᵐbu- lumɓaw	
PMofu	⁺tsufwaɗ						
PMbuko	⁺dɨsuwɗaw						
PHurzo		⁺tahkay					
PMarwa		⁺gwaltir Bald zàá.ɓə̀là					
PUldeme						⁺ʔalɓit	
Sukur				⁺mɨli			
Daba	⁺dza[maˀ]fad	⁺dɨrfatakan					
Bata	⁺tambiɗaw						⁺ʔilǮiŋ
Gidar-Musgu							
PGidar		⁺vaytak					
PMusgu		⁺tikɨla					⁺ǯiyaŋay
Muskum	⁺hɨrayfudɨ						
Kotoko							
PNKotoko		⁺inu tay finger 1 ⁺taya inaw 1 finger					
PCKotoko		⁺inu tɨku finger 1					
PIsland Kotoko							
PSKotoko		⁺tsiya … (minus 1)					
Masa							
PSMasa	⁺tifardɨʔaw						
PNMasa							⁺ǯiyaŋa

	10.1	10.2	10.3	10.4	10.5	10.6	10.7
Tera							
PTera	⁺gwaŋ						
PHwona		*kʷɨm					
Bura-Margi		*kʷɨma					
Bana	*gʷɨm						
Lamang-Mandara							
PLamang	*ɣʷaŋ						
PMandara				*kɨlawa			
PPodokwo			⁺dʒim				
Mafa-Mada							
PMuktele				⁺kulaw			

PZelgwa				⁺*kuraw*	
PMafa				**kʷiraw*	
PMefele			⁺*dimbaku*		
PMada			⁺*dzumik̫*		
PMofu				**kiraw*	
PMbuko				⁺*kuraw*	
PHurzo			⁺*dzim*		
PMarwa				**kiriⁱ̫*	
PUldeme				⁺*kʷilaw*	
Sukur	**waŋ*				
Daba	⁺*wam*			⁺*guɓ*	
Bata				⁺*[g]paw*	
Gidar-Musgu					
PGidar				⁺*kilaw*	
PMusgu					⁺*dogo*
Muskum	*gum*Röder			⁺*kuru*	
Kotoko					
PNKotoko		**kan*			
PCKotoko		**hikan*			
PIsland Kotoko		**hikan*			
PSKotoko					⁺*liɓaku*
Masa					
PSMasa				**gup*	
PNMasa					⁺*dogo*
Notes	cf. WCh ⁺*guma* 10	cf. ECh: Mawa *kʷaːyan* Mokulu *kòomá* 10	cf. Adm: Mundang **ʒémà** 10	CCh **kiriw*; cf. ECh ⁺*kuraw*; WCh: Ron: Bokkos *hừrè* 10; ECu **kurḍan-*	cf. CSud ⁺*kpuu* 10	cf. ECh: Sarwa *doko* 10; CSud ⁺*dogo* 10

Note: The symbol ⁺ indicates the preliminary reconstructions representing an intuitive 'common denominator', while the traditional asterisk * is used for standard reconstructions based on regular sound correspondences; in the case of Central Chadic these reconstructions follow Richard Gravina.

Abbreviations: Adm Adamawa; Arb Arbore; B Bole; Bald Baldamu; Bc Bauchi; C Central; Ch Chadic; Cu Cushitic; E East; Gw Gwandara; Hs Hausa; Jrw Jarawan; Kan Kanuri; M-M Mafa-Mada; N North; NS Nilo--Saharan; Om Omotic; P Proto-; S South; Sah Saharan; Sem Semitic; Sud Sudanic; Tg Tangale; W West; Zagh Zaghawa.

Conclusion

Taking in account the external cognates in other Afroasiatic branches (see Blažek 1991, 1999, 2017, 2018a, 2018b), the following Central Chadic numerals of the first decad should be determined as inherited:

§1.4. CCh ⁺tɨk-aw/-ay; besides §1.12 PMafa ⁺sitad; §1.13. PMefele ⁺(ʔa)mita;

§2.2. CCh *sɨwra < PCh ⁺ciraw;

§3.1. CCh *hɨkin < PCh ⁺kanu;

§4.1. CCh *wɨpadˀ < PCh ⁺fwadî < ⁺far(i)du;

§5.3. CCh ⁺ŝemi(S);

10.4. CCh *kɨrɨw < PCh ⁺kuraw.

The numerals '6', '7', '8', '9' are analyzable as results of various arithmetic operations:

6 = 3 + 3 ... §6.1. This pattern looks as the most frequent one for the numeral '6'.

6 = 2 x 3 ... §6.2.

6 = (5) + 1 or 5 + (1) ... §6.3.

7 = 3 + 4 ... §7.1.

7 = (5) + 2 ... §7.2.

8 = 4 + 4 ... §8.1. This pattern looks as the most frequent one for the numeral '8'.

8 = ? + 4 or ? x 4 ... §8.2.

8 = (10) − 2 ... §8.4.

8 = (5) + 3 ... §8.5.

9 = (5) + 4 or 4 + 5 ... §9.1.

9 = (10) − 1 ... §9.2.

Acknowledgement

References

Barreteau, Daniel & Le Bléis, Yves. 1990. *Lexique Mafa: Langue de la famille tchadique parlée au Cameroun*. Paris: ORSTOM.

Barreteau, Daniel & Jungraithmayr, Hermann. 1993. Calculs lexicostatistiques et glottochronologiques sur les langues tchadiques. In Barreteau, Daniel & von Graffenried, Charlotte (eds.), *Datation et chronologie dans le bassin du lac Tchad*, 103-140. Paris: Éditions de l'Orstom.

Barth, Heinrich. 1862. *Sammlung und Bearbeitung Central-Afrikanischer Vokabularien*, Band 1: *Umfassendere Vokabularien der Kanúri-, Tédā-, Hausa-, Fulfúlde-, Soṅγai-, Lógonē-, Wándalā-, Bágrimma- und Māba-Sprachen: Einleitung,* Kap. 1-6, Fürwörter, Partikeln, Zahlwörter, Zeitwörter. Gotha: Perthes.

Barth, Heinrich. 1912[1968]. Select comparative vocabulary of twenty-four Central African idioms (Ms. 1852). In Benton P.A. (ed.), *The languages and peoples of Bornu*, vol. I: *Notes on some languages of the Western Sudan*, 77-144. London: Cass.

Blažek, Václav. 1991. A comparative approach to Afrasian numerals. In Mukarovsky, H.G. (ed.), *Proceedings of the Fifth International Hamito-Semitic Congress* (Vienna, Sept – Oct 1987), vol. I: *Hamito-Semitic, Berber, Chadic*, 29-44. Wien: Afro-Pub.

Blažek, Václav. 1999. *Numerals: Comparative-etymological analyses of numeral systems and their implications* (Spisy Masarykovy university v Brně 321). Brno: Filosofická fakulta.

Blažek, Václav. 2017. Omotic Numerals. *Folia Orientalia* 54, 2017, 63-86.

Blažek, Václav. 2018. The numerals of West Chadic. In Ibriszimow, Dymitr (ed.), *Topics in Chadic linguistics IX: Papers from the 8th Biennial International Colloquium on the Chadic Languages (Bayreuth, February 4-5, 2016)*, 31-46. Köln: Köppe.

Blažek, Václav. 2018a. Cushitic numerals. *Folia Orientalia* 55, 33-60.

Blažek, Václav. 2018b. East Chadic numerals in comparative perspective. In McIntyre, Joseph A. & Schreiber, Henning (eds.), *Topics in Chadic linguistics VIII: Papers from the 7th International Colloquium on the Chadic Languages (Hamburg, 12-14 September 2013)* (Chadic Linguistics 12), 11-36. Köln: Köppe.

Blench, Roger. 2009a. *Benue-Congo wordlist*. Ms.
(http://www.rogerblench.info/Language/Niger-Congo/BC/General/BCCW%20composite.pdf)

Blench, Roger. 2009b. *Baka: A highly endangered language of Northern Cameroun*. (http://www.rogerblench.info/Language/Afroasiatic/Chadic/Central/Baka/Baka%20wordlist%20Cameroun%202009.pdf)

Blench, Roger. 2010. *A dictionary of Bura*.
(http://www.rogerblench.info/Language/Afroasiatic/Chadic/Central/Bura/Bura%20dictionary%202010.pdf)

Blench, Roger. 2012. *Dictionary of Lis ma Run (Ron of Bokkos)*.
(http://www.rogerblench.info/Language/Afroasiatic/Chadic/West/Ron/Lis%20ma%20Run%20Dictionary.pdf)

Blench, Roger & Dendo, Malam. 2006. *Jarawan Bantu: New data and its relation to Bantu*. Ms.
(http://www.rogerblench.info/Language/Niger-Congo/Bantu/A60/Jarawan/Jarawan%20Bantu.pdf)

Caron, Bernard. 2001. Guus, aka Sigidi (Chadic, west-B, South-Bachi): Grammatical notes and vocabulary. *Afrika und Übersee* 84. 1-60.

Caron, Bernard. 2002. Dott, aka Zoɗi (Chadic, West-B, South-Bachi): Grammatical notes, vocabulary and text. *Afrika und Übersee* 85. 161-248.

Caron, Bernard. 2005. *Za:r: Dictionary, grammar, texts*. Ibadan (Nigeria): IFRA (Institut Français de Recherche en Afrique). (https://shs.hal.science/halshs-00644492/file/Caron2005_Zaar.pdf)

CDAS – see Takács 2004.

de Colombel, Véronique. 1997. *La langue ouldémé*. Paris: Les Documents de Linguistique Africaine.

Cosper, Ronald. 1999. *Barawa lexicon: A wordlist of eight South Bauchi (West Chadic) languages: Boghom, Buli, Dott, Geji, Jimi, Polci, Sayanci and Zul*. Munich: Lincom Europa.

Gaudefroy-Demombynes, Maurice. 1907. Documents sur les langues de l'Oubangui-Chari. In *Actes du 14e congres international des orientalistes* 2, 172-330. Paris: Leroux. (https://gallica.bnf.fr/ark:/12148/bpt6k65817789/f15.image.texteImage)

Gerhardt, Ludwig. 1982. Jarawan Bantu: The mistaken identity of the Bantu who turned north. *Afrika und Übersee* 65. 75-95.

Gravina, Richard. 2014. *The phonology of Proto-Central Chadic: The reconstruction of the phonology and lexicon of Proto-Central Chadic, and the linguistic history of the Central Chadic languages*. Utrecht: Landelijke Onderzoekschool Taalwetenschap (Netherlands Graduate School of Linguistics).

Gravina, Richard. n.d. *Central Chadic reconstructions*. (http://centralchadic.webonary.org/)

Jaggar, Philip J. 1988. Guruntum (gùrdùŋ) (West Chadic - B): Linguistic Notes and Wordlist. *African Languages and Cultures* 1/2. 169-189.

Jungraithmayr, Herrmann. 1964. Materialien zur Kenntnis des Chip, Montol, Gerka und Burrum (Südplateau, Nordnigerien). *Afrika und Übersee* 48. 161-181.

Jungraithmayr, Herrmann. 1970. *Die Ron-Sprachen: Tschadohamitische Studien in Nordnigerien*. Glückstadt: Augustin.

Jungraithmayr, Herrmann & Ibriszimow, Dymitr. 1994. *Chadic Lexical Roots*. 2 vols. Berlin: Reimer.

Klaproth, H. Julius. 1826. Vocabulaire du Mandara. In *Essai sur la langue du Bornou, suivi des vocabulaires du Begharmi, du Mandara et de Timbouctoo*, 34-36. Paris: Impr. de Decourchant.

Kraft, Charles H. 1981. *Chadic wordlists*, vols. I-III. Berlin: Reimer.

Lebeuf, Jean-Paul. 1942. Vocabulaires kotoko: Makari. Goulfeil, Kousseri, Afadé. *Bulletin de l'Institut Français d'Afrique noire* 4. 160-172.

Lukas, Johannes. 1936. Die Logone-Sprache im Zentralen Sudan. *Abhandlungen für die Kunde des Morgenlandes* 21(6). Leipzig.

Lukas, Johannes. 1937. *Zentralsudanische Studien:* Wörterverzeichnisse der Deutschen Zentral-Afrika-Expedition 1910/11, *nachgelassene Aufnahmen von Gustav Nachtigal und eigene Sammlungen*. Hamburg: Friedrichsen, de Gruyter & Co.

Lukas, Johannes. 1939. Die Sprache der Buduma im zentralen Sudan: Auf Grund eigener Studien und des Nachlasses von G. Nachtigal. *Abhandlungen für die Kunde des Morgenlandes* 24(2). Leipzig.

Lukas, Johannes. 1941. *Deutsche Quellem zur Sprache der Musgu in Kamerun*. Berlin: Reimer – Hamburg: Friedrichsen, de Gruyter (Beihefte zur *Zeitschrift für Eingeborenen-Sprachen* 24).

Lukas, Johannes. 1965. Das Hitkalanci, eine Sprache um Gwoza (Nordostnigerien). *Afrika und Übersee* 48. 81-113.

Lukas, Johannes. 1970. *Studien zur Sprache der Gisiga (Nordkamerun)* (Afrikanistische Forschungen 4). Hamburg – Glückstadt: Augustin.

Meek, Charles K. 1931. *Tribal studies in Northern Nigeria*. London: Kegan Paul, Trench, Trubner.

Migeod, Frederick W.H. 1922. Ngala, and its Dead Language. *The Journal of The Royal Anthropological Institute of Great Britain and Ireland* 52. 230-241.

Mouchet, Jean. 1950. Vocabulaires comparatifs de quinze parlers du Nord-Cameroun. Études Camerounaises 3(29-30). 5-74.

Mouchet, Jean. 1953. Vocabulaires comparatifs de sept parlers du Nord-Cameroun. Études Camerounaises 6(41-42). 136-199.

Newman, Paul. 1964. A word list of Tera. *The Journal of West African Languages* 1(2). 33-50.

Newman, Paul. 1965. A brief note on the Maha language. *The Journal of West African Languages* 2. 57-58.

Newman, Paul. 2013. *The Chadic language family: Classification and name index*. (http://lah.soas.ac.uk/projects/megachad/misc.html)

Pozdniakov, Konstantin. 2018. *The numeral system of Proto-Niger-Congo: A step-by-step reconstruction* (Niger-Congo Comparative Studies 2). Berlin: Language Science Press.

Rossing, Melvin O. 1978. *Mafa-Mada: A comparative study of Chadic languages in North Cameroun*. Madison (Wisconsin). (Doctoral dissertation.)

Sachnine, Michka. 1982. *Dictionnaire lamé-français / Lexique français-lamé*. Paris: L'Agence de coopération culturelle et technique.

Schuh, Russell G. 1978. *Bole-Tangale languages of the Bauchi area (Northern Nigeria)*. Berlin: Reimer.

Schuh, Russell G. 1984. West Chadic vowel correspondences. In Bynon, James (ed.), *Current progress in Afro-Asiatic Linguistics: Papers of the Third International Hamito-Semitic Congress (London, 1978)* (Current Issues in Linguistic Theory 28), 167-223. Amsterdam: Benjamins.

Schuh, Russell G. 2001. Shira, Teshena, Auyo: Hausa's (former) eastern neighbors. *Sprache und Geschichte in Afrika* 16/17, 387-435.

Schuh, Russell G. & Shalanguwa, Elisha. 1983. *Bura-English-Hausa dictionary*. Ms.
(http://www.rogerblench.info/Language/Afroasiatic/Chadic/Central/Bura/Schuh%20Bura%20dictionary%20Unicode.pdf)

Seignobos, Christian & Tourneux, Henry. 1984. Note sur les Baldamu et les langue (Nord-Cameroun). *Africana marburgensia* 17(1). 13-30.

Shimizu, Kiyoshi. 1978. *The Southern Bauchi group of Chadic languages*. Marburg: *Africana Marburgensia*, Special Issue 2.

Shimizu, Kiyoshi. 1982. Die Nord-Jos-Gruppe der Plateau-Sprachen Nigerias. *Afrika und Übersee* 65. 161-210.

Shryock, Aaron. 1997. The classification of the Masa group of languages. *Studies in African Linguistics* 26(1). 29-62.

Skinner, Neil. 1977. North Bauchi Chadic languages: Common roots. *Afroasiatic Linguistics* 4(1). 1-49.

Skinner, Neil. 1996. *Hausa Comparative dictionary*. Köln: Köppe.

Sölken, Heinz. 1967. *Seetzens Áffadéh: Ein Beitrag zur Kotoko-Sprachdokumentation*. Berlin: Akademie-Verlag.

Starostin, George. 2010. Glottochronological classification of Afroasiatic languages. Ms.

Stolbova, Oľga V. 1987. Sravniteľno-istoričeskaja fonetika i slovar' zapadnočadskix jazykov. In Porxomovskij, V.Ja. (ed.), *Afrikanskoe istoričeskoe jazykoznanie*, 30-293. Moskva: Nauka.

Stolbova, Olga V. 1996. *Studies in Chadic comparative phonology*. Moscow: Diaphragma.

Stolbova, Olga V. 2005. *Chadic lexical database*. Issue I: *L, N, NY, R. Kaluga: Poligrafiya.*

Stolbova, Olga. 2009. *Chadic lexical database*. Issue III: *Sibilants and sibilant affricates*. Moskva: Institute of Oriental Studies of Russian Academy of Sciences.

Stolbova, Olga. 2011. *Chadic lexical database*. Issue IV: *Velars*. Moskva: Institute of Oriental Studies of Russian Academy of Sciences.

Strümpell, Kurt. 1910. Vergleichendes Wörterverzeichnis der Heidensprachen Adamauas. *Zeitschrift für Ethnologie* 42. 444-470.

Strümpell, Kurt. 1923. Wörterverzeichnis der Heidensprachen des Mandara-Gebirges (Adamaua). *Zeitschrift für Eingeborenen-Sprachen* 13. 47-75, 109-149.

Takács, Gábor. 2004. *Comparative dictionary of the Angas-Sura languages*. Berlin: Reimer.

Tourneux, Henry. 1977. Une langue disparue: Le mùskûm. *Africana Marburgensia* 10(2). 13-34.

Tourneux, Henry. 1982. *Lexique pratique du Munjuk des rizières: Dialecte de Pouss*. Paris: Geuthner – Orstom.

Tourneux, Henry. 1990. Place du masa dans la famille tchadique. In Mukarovsky, Hans (ed.), *Proceedings of the Fifth International Congress* (Vienna, Sept 1987), vol. 1, 249-260. Wien: Afro-Pub.

Tourneux, Henry. 2005. La numération dans les parlers 'kotoko' du Cameroun. In Mengozzi, Alessandro (ed.), *XI Incontro italiano di linguistica camitosemitica* (Bergame 2003), 113-124. Milano: F. Angeli.

Wolff, Ekkehard. 1971. Die sprachliche Situation im Gwoza-Distrikt (Nordostnigeria). *Journal of African Languages* 10(1). 61-74.

The Polish Catholic Church on the life of Poles at the beginning of the millennium: A corpus-assisted Critical Discourse Analysis

Victoria Kamasa

Adam Mickiewicz University in Poznań
vkamasa@amu.edu.pl | ORCID: 0000-0002-8975-9456

Abstract: The Catholic Church plays a vital role in Polish public life, strengthened only by the political change, which brought to power the right-wing conservative Law and Justice party. This role could be seen as a reflection of Poland's religious landscape. Yet, it is accepted only in restricted domains of life: while 61% of people accept the Church taking the stance in matters of morality, 55% disapprove of the Church voicing opinions about the Parliament's bills (CBOS 2013).

This study aims to analyse which of these domains are represented in the corpus of official documents of the Polish Catholic Church (241 000 words, 259 documents).To achieve this aim, the study, situated within the corpus-supported Critical Discourse Analysis paradigm (van Dijk 1993, Baker 2006), combines collocation and concordance analysis methods.

The findings reveal the presence of both religious and secular dimensions, although in different manners. The former has a rather individual character, whereas the latter concerns public life. Overall, the results support the view of the Polish Catholic Church as an administrator of values (Graff 2010) and point to the great self-involvement of the Church.

Keywords: Catholic Church in Poland, religion in Poland, Polish Catholicism, Critical Discourse Analysis, Corpus-Assised Discourse Studies

Introduction

Mounting research points to the central role of the Catholic Church in political and social life in Poland (e.g. Graff 2010, Heinen & Portet 2010, Mandes 2020). The vast majority of Poles declare to be Catholic as there is an inherent relationship between national identity and Catholicism, i.e. the Church played an important role in the struggle for Poland's independence and (as a result) has considerable influence on the political scene. This is merely a fraction of the arguments confirming the social significance of the Catholic Church. However, arguments to the contrary have recently started to emerge

and take shape. These include the permanent social disapproval of political involvement on the part of Church officials, accompanied by the belief that the far-reaching influence on political life by this social group is unacceptable (Grabowska 2022b). Moreover, the vast majority of Poles perceive religion to be an aspect of their lives to be confined to the private sphere. This gives rise to the question how the institutionalised Polish Catholic Church positions itself toward Poles' lives. Does it restrict itself in its statements to guidelines concerning their religious and spiritual life? Does it take on the role of an authority in matters concerning the private lives of its believers? If speaking on social and political issues, what discursive identity does the Catholic church assume: a commentator explaining to the faithful the current social and political life from the perspective of faith and its dogmas? A critic who points out actions inconsistent with the morality adopted within the Catholic doctrine? An authority whose right is to require and impose certain actions in the secular domain? These broad questions have inspired the corpus-assisted analysis of official documents of the Polish Bishops' Conference presented in this paper.

Social and political role of the Polish Church

Several metrics of religiousness indicate the fundamental role of Catholicism in Poland. According to the census from 2011, 96 per cent of people who answered the question about their religion chose the Roman Catholic faith (88 per cent of the total population) (GUS 2011). This trend is also confirmed by more recent data collected by the Public Opinion Research Centre (Centrum Badania Opinii Społecznej, CBOS): 'Our systematic research shows that since the end of the nineties, more than 90 per cent of those surveyed (92 - 97 per cent) have invariably considered themselves as believers'[1] (Boguszewski 2017a) and mirrored in results of the International Social Survey Programme ISSP 2008 – Religion III 2008 (86 per cent) (ISSP Research Group 2018). Also, data on participation in services reflects such declarations. According to ISSP, nearly 35 per cent of Poles attend a service at least once a week (ISSP Research Group 2018),. Also, data on participation in services reflects such declarations. According to data collected by the CBOS, nearly 40 per cent of Poles attend a service at least once a week (CBOS 2022), while observations carried out by the Institute of Catholic Church Statistics (Instytut Statystyki Kościoła Katolickiego, ISKK) put this percentage at the approximate level of 37 per cent (ISKK 2020). This puts Poland among other traditionally Catholic countries in Europe such as Irland (87 per cent Catholics, 44 per cent attending at least once a week), Italy (89 per cent Catholics, 22 per cent attending) and Spain (75 per cent Catholics, 20 per cent attending) (ISSP Research Group, 2018). Life's landmark events' celebration also take the Catholic form: 98 per cent of children born in 2016 were baptised, 62 per cent of weddings in Poland were held in the Catholic Church[2]. As a result, Catholicism seems to be widespread in Poland at the level of declarations

---

[1] All translations from Polish publications and other Polish texts are provided by the author.

[2] Due to the Concordat from 1993 Church wedding is recognized by the state.

as well as religious practices, both in everyday life and on festive occasions. The commonness of declarations of faith may attribute the Church the legitimacy for statements about the lives of all Poles, not excluding ones outside this religious community. Furthermore, the commonness of religious practices can sanction statements concerning various spheres of their lives, not limited to spiritual or religious life.

The important role of Catholicism in the life of Poles is also evidenced by its connection with national identity. Over 80 per cent of Poles positively assess the impact of Polish baptism on both the Polish statehood and the Polish culture (Boguszewski 2016). This percentage varies only slightly throughout the political spectrum; therefore, we may assume that most Poles are satisfied with the fact that they live in a Catholic country. This is also reflected in the discursive construction of the Pole-Catholic identity. In communist times, it was one of the most important parts of the Catholic Church's self-characteristics (Kominek 2003). As Kominek (2003) puts it: "The result of this permanent connection [between nation and Catholicism, VK] was in the past the formula of the Pole-Catholic, which almost put these two concepts on a par with each other". The political transformation has not brought about any significant changes in this respect. Skowronek (2006) shows a similar connection between being Polish and being Catholic within a dataset of contemporary pastoral letters. Moreover, Makuchowska (2011) draws our attention to the strengthening of the image of non-Catholics as enemies of the nation, Christianity, and the Church after the political transformation of 1989. This clear discursive connection between national and religious identities may lead to discursive statements concerning the life of the nation, i.e. not only the private life of the faithful but also – more broadly – the social or political life in Poland.

The critical role of the Catholic Church in the post-war history of Poland is hardly questionable. During the communist era (1945-1989) its crucial position was recognised by all important social actors: the ruling communist party and the opposition. Moreover, the Church was well aware of this. Fighting against this powerful institution became one of the most important goals of the communist party, which was carried out by various means, beginning with anti-church propaganda, through surveillance of the clergy and Catholic activists, closing Catholic schools, to the detention and imprisonment of priests and monks. One of the most dramatic manifestations of the party's struggle with the Church was the arrest and detention of the Primate of Poland, Cardinal Stefan Wyszyński in 1955, and the brutal assassination of Father Jerzy Popiełuszko by Security Service agents in 1984 (Skowronek 2006). What is more, for members of the opposition, churches were often a place of recluse from the oppressive communist reality. It was there that they had the opportunity to express their anti-communist sentiments freely. Masses often turned into demonstrations of patriotism and national solidarity (Skowronek 2006). The analysis of Solidarity movement documents carried out by Kominek (2003) also presents the discursive dimension of this connection between Catholicism and patriotism: in these documents the Catholic Church and the Polish nation are seen as an undivided unity. Such historical conditions may reinforce the aforementioned tendency for the Church representatives to comment on social and political life in Poland, as well as indicate that they adopt the position of active authorities or critics rather than passive observers.

The political transformation of 1989 fundamentally changed the situation of the Catholic Church in Poland. On the one hand, it was allowed to organize public religious events on a large scale; it was also granted unrestricted symbolic rights to take a stand on social and political issues. On the other hand, the freedom of speech, which was part of the democratic transformation, made the Church a subject of open criticism (Kominek 2003). According to many researchers, these changes have not led to a weakening of the role of the Catholic Church in social and political life in Poland but even strengthened it. The Church has become present in the public space. This can be exemplified by, for instance, hanging Christian crosses in public places (including the Parliament), the presence of the Church's officials during the celebration of all important public holidays, religion taught in state-run schools and financed from the state budget or, finally, the inclusion of Christian values in the constitution (Graff 2010). The institution's political influence has also increased significantly: Heinen indicates that many political decisions are motivated by the politicians' fear of losing the support of, what they consider to be, the most numerous social group, the Catholics (Heinen & Portet 2010). Borowik (2002) describes the attempts of the Catholic Church to influence the results of elections and Graff (2003) suggests the debate on antiabortion law as an example of "the power of the Church to upset the democratic process itself, and the willingness of the political elite to give in to the pressure of clergy" (Graff 2003: 112). The scope of the Church's political impact is also reflected in legal solutions adopted in Poland, such as the criminalisation of abortion (Zielinska 2000; Owczarzak 2009) against the will of the majority of voters (Graff 2003), severe restrictions on family planning services (Mishtal & Dannefer 2010) or the Conscience Clause law allowing medical practitioners to refuse health services because of conscience-based objections (Mishtal 2009). The influence of the Church on education in Poland is also visible, among others, through religious education as part of the curriculum in Polish schools (Heinen 2010) and influence on the content of 'Preparation for Life in a Family' courses (Mishtal & Dannefer 2010; Heinen & Portet 2010; Pakuła et al. 2015).

Religion as a private matter

High religious indicators are not accompanied by the acceptance of the Church's involvement in social and especially political life. The lack of acceptance remains as constant as declarations of faith or participation in religious practices. As early as by the end of the 1980s, more than half of Poles did not accept the Church's representatives' comments on political issues (Mariański 1993). The political transformation did not change this situation. At the beginning of the 1990s, the majority of Poles (61 per cent) supported the complete separation of the state and the Church (Mariański 1993), and nearly half declared that the Church should not influence socially important decisions concerning areas other than faith and religion (Mariański 1993). This lack of acceptance is even more evident when the Church attempts to exert direct influence on elections. In 2013, more than 4/5 (82 per cent) of Poles felt outraged by priests actively and publicly (also during services) supporting specific political parties. This percentage has also stayed

unchanged since the beginning of this millennium (Grabowska 2022b). The reluctance to listen to the Church's opinion on electoral issues is also reflected in the sphere of individual decisions: in the late 1990s, more than half of Poles (58 per cent) considered that a Catholic could vote for a non-believer candidate (Adamczuk 2001), while in 1995 and again in 2000, Aleksander Kwaśniewski openly declaring himself a non-believer was elected president. Given these statistical data, one can reasonably expect the Church to focus on the private sphere in the discursive construction of life.

Over the last two decades, the Church's involvement in political life in Poland has consistently been assessed as direct or even unacceptable. Thus, at the beginning of the 1990s, nearly two-thirds of Poles (64 per cent) noticed the influence of the Church on politics and politicians, assessing it as significant or decisive (Mariański 1993). The next decade does not substantially change this assessment: in 2001, two-thirds of Poles (66 per cent) still noticed the involvement of the Church in politics (Adamczuk 2001). Moreover, as early as in the early 1990s, more than 2/3 of Poles (67 per cent) considered the influence of the Church on public life in Poland as too far-reaching (Mariański 1993). The next decade witnessed only a slight change in this assessment since it decreased to 60 per cent in 2006 (Mariański 2011), and in 2007, nearly half of Poles (49.6 per cent) indicated that the influence of the Church on the life of the country should be less significant (Mariański 2011). Moreover, the political involvement of the Church is perceived not only as something uncalled for but even as one of the fundamental issues internal to the Church in Poland: in 2019, over 1/3 of Poles identified the involvement of the Church in politics as one of the main problems of the Church. It takes second place after the uncovered cases of paedophilia indicated by 60 per cent of the respondents (Głowacki 2019). Such a negative assessment of the Church's involvement in politics might suggest that its discourse will tend to avoid statements about social and, in particular, political life.

Not only do Poles oppose the Church to being involved in political life, but they also treat religion as a private sphere of their lives, reserved only for themselves or for the people closest to them. For example, conversations on religious matters are held only in the closest family circles: 43 per cent of Poles declare that they avoid such conversations at work or in educational settings, and 27 per cent even in conversations with friends (Zaręba 2010). Putting religion in the private sphere is also indicated by the belief that education in a state-run school should be based only on norms and values stemming from general humanism relying on common principles, and recognised both by believers and non-believers. Only one in six believes that it should be based on the ethics of the Catholic Church (Boguszewski 2017b). The privacy of religion among Polish Catholics is also reflected in the separation of moral norms from religiousness. Nearly 70 per cent of respondents are of the opinion that what good and evil should be primarily a personal matter for every human being, and only 15 per cent believe that this should be primarily dictated by God's laws (Boguszewski 2017b). As many as 1/3 of Poles do not accept the Church's statements on moral issues (Grabowska 2013). Catholic sexual morality, of which a large number of Catholics remain critical, seems to be a particularly sensitive area. At the same time, the Church's interference in the sphere of sexuality is considered to be overly extensive and intrusive (Baniak 2007). Such views of Polish

Catholics suggest that the discourse of the Catholic Church might focus only on the religious life of its followers.

The tension arising from the discussed research concerns the functioning of the Church in the social sphere. There is also a clear variation in the level of social acceptance of this role, due to the domains concerned. Moreover, there is no doubt that the Catholic Church is present in the religious life of Poles, as evidenced by the high rates of religious practice. However, its involvement in the private lives of Poles is perceived more ambiguously. On the one hand, they commemorate the most important events of their lives through religious celebrations (e.g. weddings). On the other hand, though, they object when the Church interferes in the sphere of their morality or sexual life. When it comes to social life, we are dealing with the acceptance of the broad presence of the Church in public life and a firm disagreement with its involvement in politics. From the perspective of this ambivalence, the question about the role of the Church per se in relation to these three domains is interesting. In this article, I will therefore focus on addressing the following research questions:

1. How does the Catholic Church position itself discursively in relation to the religious life of Poles?
2. How does the Catholic Church position itself discursively in relation to the private life of Poles?
3. How does the Catholic Church position itself discursively in relation to issues of social and political life in Poland?

These questions are approached from the perspective of Critical Discourse Analysis (CDA), understood as "a form of academic inquiry which aims to achieve a better understanding of how societies work" (Baker & McEnery 2005). I consider the institutional Polish Catholic Church as a powerful agent within society. I aim to explore how its discourse legitimizes and reinforces this power and to understand in which domains of life this process occurs.

Methods and materials

Corpus

The corpus analysed in this study comprises all documents published on the website of the Polish Bishops' Conference (Konferencja Episkopatu Polski, KEP). The choice of documents from this institution is based on the belief that they most accurately represent the official discourse of the institutional Catholic Church in Poland. The Polish Episcopal Commission is "a permanent institution established by the Holy See (..)." (KEP 2009), and its tasks include overseeing the religious and moral life of Catholics in the state, as well as maintaining proper relations between the Church and the State (KEP 2009). I, therefore, recognise this institution as representative of the Catholic Church in Poland as a whole and also speaks on its behalf. Such an approach to the selection of data can be defined, as Bednarek (2008) suggests, as a semi-automatic analysis of medium-size corpora, which limits the size of the corpus to a volume that allows a manual analysis

of its selected aspects. This permits the use of qualitative techniques and therefore a more complete reconstruction of the discourses under investigation.

The corpus analysed in this study is composed of all texts published between 02.05.1998 and 01.03.2013 in the 'Documents' section of the official website of the Polish Episcopal Conference (www.episkopat.pl). These documents were collected by means of Teleport Pro ver. 1.64, which downloaded all subpages with the address starting from www.episkopat.pl/dokumenty/. I then manually extracted the texts of individual documents from these downloaded documents. They were also manually cleared of their HTML codes and converted into .txt format and UTF-8 encoding files. This resulted in a corpus of 241 thousand words made up of 259 documents. Finally, the corpus was lemmatised by the use of PSI-Toolkit tools (Graliński et al. 2013) resulting in a corpus of 13,000 unique words.

Analysis

The analysis was conducted in the following steps (described in detail below):
1. Identification of collocation for the term *life* (Polish: *życie*).
2. Analysis of semantic preference for the term *life*.
3. Analysis of concordances in a selected group of collocations:
    a. Classification of concordance into one of two domains: religious or secular.
    b. Classification of concordance due to communication activity: criticising, instructing, or describing.
    c. Analysis of quantitative relations between domains and communication activities.
    d. Qualitative analysis of concordance for selected types of communication activities.

In the first stage, collocations for the term *life* were identified. Here, collocation is understood as "more frequent than random co-occurrence of two words within a predefined range" (Gabrielatos & Baker 2008). Discourse analysts assume unconscious interference of diverse meanings related to the cumulative effect of previous contacts with a given word or a collocation (Gabrielatos & Baker 2008) and therefore use collocation analysis as a starting point for further, more in-depth analyses of corpus data (cf. e.g. Freake et al. 2010, Forchtner & Kølvraa 2012). This choice was dictated by the richness of the material studied (1176 occurrences of the term *life*), which required analysis at a more general level, allowing me to see general patterns, which led to more nuanced analyses. Collocations were identified using the AntConc program and MI statistics (Anthony 2014). The list of collocations was generated within the window of three words to the right and left of the term *life*. Statistical relevance was assessed with the t-test ($\alpha<0.005$) and five as a minimum number of co-occurrences.

After determining the list of collocations, I carried out an analysis of semantic preference, i.e. the tendency of a specific lexical unit to frequently co-occur with a series of units belonging to one semantic field (Salama 2011). Following this, I reconstructed thematic groups and ascribed individual collocations to them. I aimed to identify groups with an average level of generality so that they could preserve the specificity of the individual collocates while avoiding large fragmentation. The classification into thematic

groups was based on the collocates themselves, and only in cases of doubt did I refer to text usage. Function words were excluded from the analysis of semantic preference. After classifying the collocates into individual semantic groups, I carried out a quantitative analysis of interrelations within the groups, as well as between them.

The next stage was focused on the analysis of concordances for collocates related to secular life. This choice was dictated by the tension concerning the involvement of the Catholic Church in secular life in Poland. I expected that an in-depth analysis of the concordances of this particular group would allow me to reconstruct the position assumed by the Catholic Church regarding the lives of Poles. The first stage of the analysis was the classification of concordances into one of the two domains: secular or religious. Singling out such domains was also due to the tension between the high level of Poles' religiousness described above and their lack of acceptance of the Church's interference in secular life. When classifying concordances, I decided that only concordances of a clear religious nature should be included in the religious domain. In the second stage, I divided the same concordances according to the type of communication activity. I distinguished three types of such activity: Criticising, instructing, and describing. In this case, too, the distinction of such domains was dictated by the tension which was clearly expressed in the opinions of Poles: from frequent religious practices and acceptance of the presence of the Church in public space to aversion to opinions expressed by the Church on non-religious topics such as social life or sexual life. When categorising specific concordances, I employed linguistic markers to identify criticism and instruction. These included:

– Criticism: references to deficiency or excess (e.g. *lack of*, *too often)*, expressions related to conveying fear (e.g. *our concern is*), and the use of negative adjectives (e.g. *unfriendly*, *unfavourable*).
– Instruction: use of imperatives, modal verbs (e.g. *must*, *may*), and verbs indicating the necessity of action (e.g. *need*, *expect*, *ought*).

I first sought criticism then instructing, and in the case of their absence, I considered a given example a description. Next, I analysed the quantitative relationships between domains and communication activities. At the final stage, the examples in which criticism or instructions were expressed were subjected to a further detailed analysis leading to the identification of the addressee of the criticism/instructing and their subject.

Results

Semantic preference

The term *life* in the examined corpus collocates with 246 words, 139 reaching the statistical significance of less than 0.005. After excluding function words (such as *and*, *in*, *to*) I retrieved a list of 73 collocates, which in total co-occurred 1588 times with the term *life*. These collocates were further analysed for semantic preference (Stubbs 2005). On this basis, I identified seven main semantic groups:

1. religion (*consecrated*, *eternal*, *society*, *priestly*),
2. secular life (*daily*, *matrimonial*, *personal*, *public*),
3. quality (*style*, *meaning*, *field*, *testimony*),
4. action (*save*, *introduce*, *defence*, *sacrifice*);
5. values (*dignity*, *value*, *truth*, *love*),
6. stage (*conception*, *death*, *day*),
7. subject (*human* (adj), *person*, *human* (n)).

All the examined collocations referred to the human being, which indicates a discursive construction of life as something exclusively concerning people. This is also confirmed by the collocates in the *subject* category. Life is also constructed as divided into stages, can be the subject of various activities from defence to sacrifice, and is also subject to evaluation in terms of its quality and with regard to certain values.

The quantitative distribution of the term *life* with collocates from each category is presented in Figure 1:
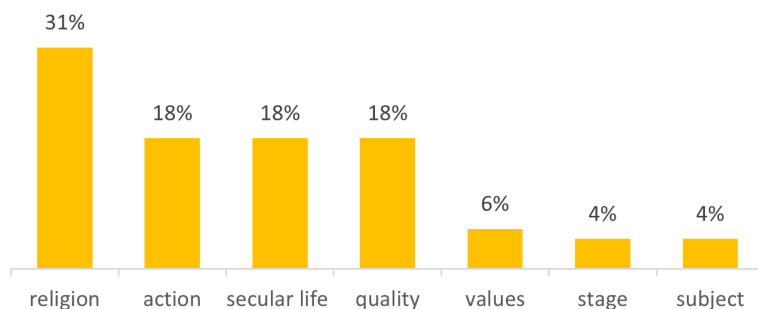

Figure 1: Semantic preference of term life

Slightly less than 1/3 of the examined cases refer to life in a religious context. This may point to the domination of the topic characteristic of the examined discourse, but it may also indicate that life is reduced to its religious character. A more comprehensive picture is provided here by dividing collocates from this category into two dimensions: institutional with examples such as *consecrated*, *association*, *priestly*, *institute*, *Church*, and spiritual (e.g. *eternal*, *vocation*, *holiness*, *spiritual*, *God*, *divine*, *faith)*. The quantitative distribution of the collocates of term *life* in these dimensions is presented in Figure 2.

Collocates from the first category can be treated as indicators of the subject-specific to the discourse under investigation, while those from the second category can be seen as shifting focus to the religious dimension of life. The relatively equal distribution between both these dimensions does not permit to point to the domination of any of these dimensions. It is, however, worth noticing that in this discourse database, one in six occurrences of the term *life* is related to the institutional Church.
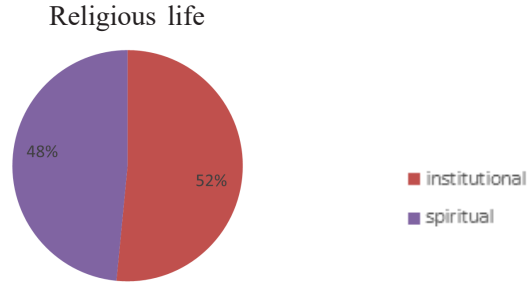
Religious life

Figure 2: Distribution of the collocates of word life in institutional and spiritual dimensions

Another group is formed by collocates concerning secular life. This group includes both collocates referring to life in a very general way, such as *everyday* or *personal*, as well as those concerning family life (*family*, *marriage*) and those indicating interest in social life (*economic*, *political*, *public*). Their quantitative distribution is shown in Figure 3.
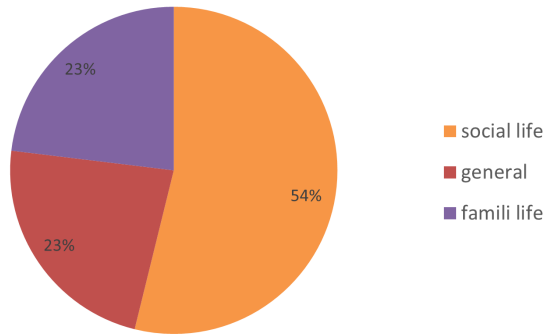

Figure 3: Distribution of collocates concerning secular life

It is worth noting that in more than half of the cases, the focus is on social or political life. This is particularly interesting given the previously discussed negative perception of the Church's involvement in the public sphere of Poles' lives. The analysis of the collocates alone, however, still has not allowed me to decide unambiguously whether it is rather explaining social life from a religious perspective or a direct involvement in this life and attempts to influence it. A clearer picture emerges from the concordance analysis for all the occurrences of term *life* with secular collocates.

In the discourse under study, secular life is mentioned both in its religious and non-religious context. In the first case, there are primarily indications of how the faithful should behave in public life e.g. *Let us be his witnesses at home, but also <u>in public life</u>. The world has the right to see Jesus' followers, who have met Him personally and live with Him every day in great intimacy*[3]. The emphasis is placed on bearing testimony to

---

[3] Due to linguistic differences, some translated examples may not include the word *life*. For the sake of clarity in the English translation, I prioritized fluency over strict word-for-word accuracy

one's faith and acting as a witness to it, which shifts the focus from the private to the public sphere.

It is striking that it is the latter, non-religious group that prevails in the official statements of the Polish Episcopate. The exact quantitative distribution of concordances is shown in Figure 4.
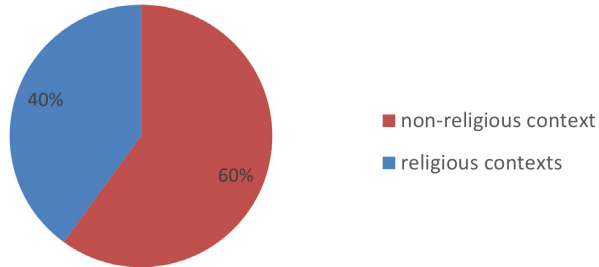

Figure 4: Distribution of contexts for secular life

In the attempts to answer the question how the Catholic Church positions itself in relation to the life of Poles, the way in which it talks about secular life is of vital importance. The passages quoted above suggest that not only does the Church describe or explain social life, but also criticises it or gives instructions concerning it. This assumption is confirmed by quantitative data (Figure 5).
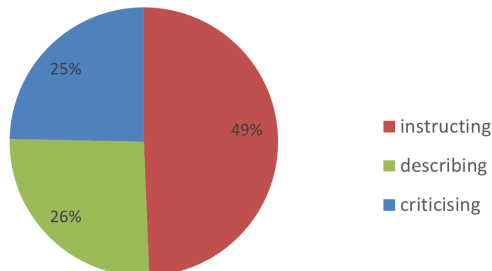

Figure 5: Distribution of communication activity

Nearly half of the statements analysed are various forms of instructions, ranging from strong imperatives (1st or 2nd person plural, e.g., *Be witnesses of Love in social life!*) to more subtle advice (e.g., *It is worth considering what areas of our personal lives we can bring more order and clarity into*). Most of these instructions neither specify who is giving them nor identify the intended audience, thereby conveying a sense of universality (e.g., *There needs to be a deep understanding of the principle of Christian obedience throughout our social life*). Therefore, in statements about secular life, the attempts to influence the shape of that life definitely prevail (75 per cent). Only every fourth statement is of a descriptive nature (e.g. *in this way, the leaven of God's grace enters the great bygone cultural, economic, and political life in the world (..)*).

In the light of this prevalence of instructing and criticising statements, the question emerges: who is the subject of this criticism, who is the intended recipient of the instruc-

tions, and what exactly is the focus or object of both the criticism and the directives. Among the addressees of the criticism, I have distinguished six groups:

1. Politicians: *We therefore share the concern about the quality and style of <u>political life</u> in our country.*
2. Society at large: *We see many worrying manifestations of <u>social life around us: egoism, greed, and rapacity</u>.*
3. We: *In our professional and personal lives, <u>we</u> have too often behaved in such a way and made decisions as if <u>we</u> doubted that love should be the most important motive for our actions.*
4. Unspecified subject (unknown *they*): *We see a clear campaign, a whole strategy for <u>eliminating God from the lives of nations, societies, and people's consciences</u>.*
5. Social phenomena such as *secularisation* and alleged aggression against the family: *Undoubtedly, a huge challenge for religious people is the galloping <u>secularisation processes</u>, which attempt to reduce people solely to the material level and to relegate religion to the margins of social life or to confine it entirely to the private sphere.*
6. *Others, such as families, young people, secularised people, you*: <u>*Our families*</u> *are experiencing confusion and doubt about the fundamental meaning of marriage and family life.*

The quantitative distribution for the individual addressees of the criticism is shown in Figure 6.
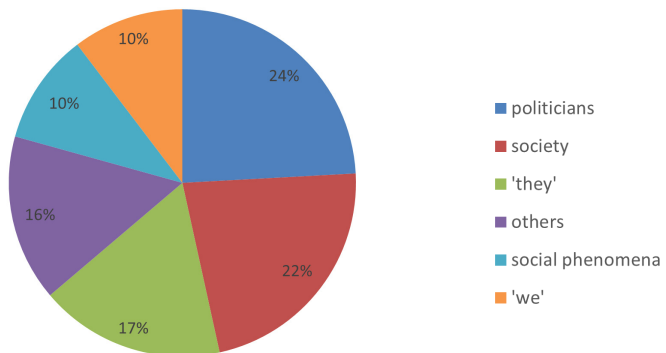
Figure 6: Distribution of criticism of different addressees

Every fourth criticism (14 out of 58) is directed at politicians. It concerns only the secular dimension of life, and it appears in statements with collocations classified as belonging to social life (the vast majority) and family life. The subjects of criticism are not directly related to the Church. Let us take the example of political culture (e.g. *<u>using the structures of public life</u>, and especially the institutions of the state <u>for individual or group benefits</u>, is a denial of properly understood politics*) or the social situation in Poland (e.g. *Given these <u>painful phenomena in our social life</u>, the question arises: are we helpless against evil?*). However, there is also criticism concerning the attitude of politicians towards the Church (e.g. *it is not uncommon for us to encounter manifestations of hostility towards Christ, His Gospel, and the Cross, as well as <u>attempts to exclude the</u>*

Church from public life). Thus, in the discourse under study, there is both the prospect of defending one's interests or position and of caring for the common good.

The next most frequent group (13 out of 58) comprises critical statements, which do not directly identify the addressee of the criticism, even though the criticism always concerns phenomena that feature a doer. In the following case, the criticism is more directly related to the religious sphere because it concerns the removal of religion from public life (e.g. W*hen anyone tries to remove Christ from public life and wants to build a Europe without Christ, it has to be said with truth and for the sake of Europe that it would be building on sand)*. Without a clear indication of those responsible, the violation of "the right to life" is also criticised (e.g. *Bishops sympathise with communities, including the media, which consistently defend the inviolable right to life, even at the price of unjust accusations and sentences)*.

As shown in Figure 6, every tenth critical statement is self-critical (6 out of 58). However, this is never a criticism of the representatives of the Church itself. The *we*-subject of these statements is always inclusive and concerns both speakers and listeners. This criticism appears in statements with collocates from the social field as well as those of a general nature. It concerns certain lifestyles (e.g. *The difficulties of everyday life meant that, although we were constantly experiencing many graces, our prayers were more often pleading than thankful)* as well as our attitude towards the family (e.g. *In social, economic, and cultural life, we allowed the family and its rights to be ignored)*. Thus, here as well, what is being criticised is not directly related to religious practices or attitudes towards the Church pointing again to the discursive involvement of the Church in social life in Poland.

The target groups of the Church's instructions are not very different from the groups which are criticised:

1. We: *let us be His witnesses at home, but also in public life.*
2. Politicians: *We are concerned that those responsible for shaping our social and political life are guided by concern for the common good (..).*
3. The family: *Above all, it encourages us to see the presence of God in the everyday life of the family (..).*
4. Other groups such as members of church movements, employers, and representatives of certain professional groups such as teachers.
5. Not specified *(..) faith should have consequences for society in the wider sense.*

There is, however, a clear difference in the quantitative distribution of the instructions (Figure 7).

Slightly more than a third of the instructions are formulated in the first-person plural (*we*). As in the case of criticism, it is an inclusive *we*, concerning both speakers and listeners (in this case, however, there are two exceptions[4]).

Among the actions that are the subject of self-instructions, many are directly linked to religion. These include following Christ (e.g. *we are called to obey His calls and to follow Him in our personal and social lives)*, witnessing to the faith with life (e.g. *we*

---

[4] (1) *they are waiting for the sacraments of which we are the ministers, (2) They await the testimony of our personal life, our faith.*
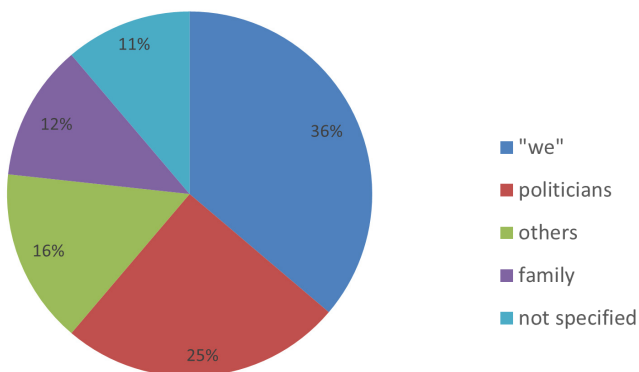
Figure 7: Distribution of instructions for different addressees

are called to be witness to the Gospel of salvation in our daily lives), with particular emphasis on witnessing in the public sphere (e.g. *Let us be His witnesses at home, but also in public life*). There are, however, also matters that are not directly related to the religious sphere, namely:

–   the protection of "the unborn life" (*On the threshold of the 21ˢᵗ century we face common tasks: protecting life, marriage, and family*),
–   specific actions related to the family (*This creates an opportunity for us to reflect on topics that are important for every Christian family, and also because of the nature of family life, important for every family*),
–   attitudes in conformity with specific values (*An important task today is to develop the imagination of mercy, which allows us to see new opportunities for the realization of grace around us*).

Politicians are the second most frequent group of instruction addressees. Every fourth instruction appearing in the studied material addresses them. These instructions take only two forms:

–   imposing standards (e.g. *The government should strengthen the position of families, and improve the level of health care for women and girls),*
–   indicating expectations (e.g. *however, we expect people and institutions responsible for the shape of political, social, and economic life to take effective action).*

It is worth noticing that in both cases these are indirect forms of instructions. The thematic scope of these instructions is also different. While self-instructions were largely linked to religious life, instructions addressed to politicians have no link to religion. They relate to three main thematic areas: the defending of "the conceived life" (e.g. *an important task is to take action to promote the right to life of all conceived children, as well as effective family policies*), ensuring an appropriate standard of living for the citizens (e.g. *we expect people and institutions responsible for the shape of political, social and economic life to take effective action so that Poles can find decent conditions for work and life above all in their homeland*), as well as to acting by a specific catalogue of principles (not necessarily of religious nature) (e.g. *according to bishops, all aims of social and political life must be achieved using always morally good means*).

Every tenth instruction touched upon the family. These instructions adopt forms similar to self-instructions, such as suggestion[5], imperative[6] and indicating the norms explicitly[7] and implicitly[8]. Thematically, however, issues related to religious life prevail here, such as the performance of certain religious practices (e.g. *Celebrating Sunday in the family may give rise to specific attitudes in the family's daily life*) or adopting a certain attitude towards God (e.g. *Above all, it encourages us to notice the presence of God in the everyday life of the family*). The only subject not related to religious life was participation in social life (*let us ask the question: how does every family participate in creating bonds of social life?*). Additionally, no guidance regarding family relationships[9] or the ways in which members should treat one another was observed.

Discussion

In my research, I sought to investigate the position that the Catholic Church in Poland assumes with regard to the religious life of Poles, their private life, as well as social and political life. Based on the results presented above, three positions can be indicated:
1. a somewhat involved observer of the religious life of Poles
2. a religious guide in private life
3. an authority deeply engaged in social and political life.

The initial mode of self-positioning by the Catholic Church becomes apparent upon a preliminary analysis of collocational patterns: one in every six collocations has a religious character and refers to spirituality, which in the official discourse of a religious organisation can be considered relatively low. The Church's positioning of a somehow involved observer was also visible in the analysis of concordances concerning secular life: less than half of them relate to the religious dimension of this life. Only slightly more than a third of the instructions refer to the community of believers (the collective we). In this case, too, the emphasis is often placed on bearing witness to one's religiousness in the public sphere. In addition, only one in ten of the criticisms examined concerns members of the Church (the collective we). This criticism also does not directly concern religious practices or spirituality.

In relation to the private life of Poles, the Church takes the position of a guide in the religious dimension of this life. The instructions addressed to the family are mostly related to widely understood religious practices. Also, the focus of statements concerning everyday life is on the religious life. The fact that all statements about private life are instructions underlines the taking of the role of a guide.

---

[5] E.g. *Above all, he encourages the recognition of the presence of God in the everyday life of the family.*

[6] E.g. *We appeal to parents: be present in the life of children and young people.*

[7] E.g. *The vocation to be wife and husband, mother and father sets absolute limits to all professional and social involvement.*

[8] E.g. *The right form of celebrating can save and restore full dignity to family life.*

[9] With one exception for the relationship between parents and children: *we appeal to parents to be present in the lives of children and young people.*

The last position taken by the Church starts to be apparent at the level of the analysis of collocations: every fifth collocate concerns secular life, and half of the collocates related to secular life concern social and political life. This trend is also confirmed by the results of concordance analysis: 60 per cent of all statements with collocates classified as social are related to the secular dimension of life. Thus, when the Church's discourse tackles social life, these statements relate to its secular dimension and are devoid of references to religion or spirituality. The interest in political life, on the other hand, is indicated by the fact that one in four of the instructions analysed and one in four of the critical statements address politicians. Given the religious nature of the texts examined, which include pastoral letters read at mass instead of homilies, such a share of social issues and those not directly related to religion or spirituality seems significant. The fact that almost half of the statements produced about it are instructions and a quarter are criticism is proof of the Church positioning itself as an authority in this domain. Thus, the vast majority (74 per cent) of statements concerning social life are made from the standpoint of a superior, a position of a subject that has the right to instruct, criticise, and impose norms. This position can also be dubbed, as Graff (2010) puts it, a manager of the values.

Each of the above-mentioned tendencies is reflected in the views of Poles, as well as their attitudes towards religion and the Church. Low interest in the spiritual life and religious dimension of secular life is visible at a similar level of interest in these aspects of religion among Poles. Most Poles have a secular approach to religious holidays and rituals and do not understand their symbolism and meaning. This applies, for example, to a completely secular way of celebrating the sacrament of marriage, a lack of understanding of the symbolism of Easter celebrations, or a lack of understanding of the meaning of Confirmation rites declared by the confirmed themselves (Baniak 2007). The low level of interest in religiousness is also evident in the relatively low acceptance of Catholic dogmas; for example, as many as 14 per cent of Catholics do not believe in the dogma of eternal damnation (Zaręba 2010), and nearly half of Poles declare that they are believers, but in their way (Boguszewski 2009). The absence of religion in the secular life of Poles is manifested, for example, by a fairly widespread acceptance of practices rejected by the Church, such as the use of contraceptives (68 per cent consider it acceptable or conditionally acceptable), premarital sex (71 per cent), permanent living with an unmarried partner (55 per cent) (Zaręba 2010) or same-sex partnership (42 per cent) (Scovil 2021). In addition, when faced with moral dilemmas, Poles prefer to be guided by their conscience rather than by the Church's guidelines (Zaręba 2010), especially in the case of married and sexual life. The prevailing opinion is that the latter sphere in particular should depend solely on the spouses themselves (Baniak 2007). Summarising his research, Zaręba (2010) highlights the tendency for references to God in everyday matters to diminish, which can be interpreted as a direct reflection of the disinterest in the spiritual and religious dimensions of secular life, as evidenced in the discourse analysed.

Poles accept the Church's position as a guide to the religious dimension of private life. This is evidenced primarily by the high rates of religious practices described above. What is more, more than 50 per cent of Poles perceive Mass as the fulfilment of their

duty to the Church (Zaręba 2010), which suggests that the Church is perceived as an institution to which the faithful are obliged to do something and which has the strength to impose this obligation. A similar way of seeing the Church and its authority is visible in the statements of parents of children who are about to have their Confirmation. They regard this sacrament as 'yet another rite to be fulfilled if "such" is the Church's wish.' (Baniak 2007).

Poles' attitude to the Church's focus on social and political life is ambivalent. On the one hand, the acceptance of the presence of the Church in the public sphere is widespread and constant: more than three-quarters of Poles accept the participation of Church representatives in state ceremonies (81 per cent), the consecration of public places, and buildings (78 per cent), the appearance of priests in public television (74 per cent) or the religious nature of the military oath (83 per cent) (Grabowska 2015). Public life is therefore considered to be a sphere to which the Church has access, while this presence outrages only a small proportion of the population (about ten per cent). Therefore, at this level, the Church's discourse remains consistent with the expectations and views of Poles. The perception of the Church's influence on political life described above also remains consistent with this trend (66 per cent of Poles see the Church's involvement in politics (Adamczuk 2001)). However, despite the discursive focus on this aspect of secular life, there is a lack of consent among Poles for the Church to interfere in political life: let us recall the data already quoted above more than half of Poles are outraged by Church representatives commenting the laws passed by the Sejm, 84 per cent do not accept that priests should tell their faithful how to vote in elections (Grabowska 2022b), and 37 per cent perceive engaging in politics as a problem faced by the Catholic Church in Poland (Głowacki 2019). Thus, while the focus on social life is not questioned and is accepted, the focus on political life is a tendency that arouses quite widespread opposition among Poles.

Conclusion

However, some reservations must be made about these results. First, the evaluation of semantic prosody was based only on collocates, without analysing the context in which they occurred. In some cases, therefore, this may have led to a different classification from the one that would have been made after having studied that context. Nevertheless, I believe that these are individual cases that do not have a significant impact on the overall picture resulting from this analysis. Secondly, in a concordance analysis, I have focused solely on the collocates categorized as concerning secular life at the expense of examining statements about life in religious or value contexts. I assume that the tension outlined in the first part concerning the functioning of the Catholic Church in Poland primarily concerns its presence in social or public life, i.e. the area of secular life. Therefore, the analysis of this very area of discourse about life allows for the most adequate reconstruction of the position that the Catholic Church in Poland takes.

These findings demonstrate how the Catholic Church in Poland constructs and sustains its authority through its discursive practices. They also reveal that the positioning of the

Church in life in Poland aligns with the social situation and is against the will of the majority of Poles. However, recent years have brought some signs of change in this social situation. The large-scale protest against harsh anti-abortion law (November 2020) had strong anti-church and anticlerical dimensions. The number of men entering the priesthood is crumbling: from 725 starting their education in seminaries in 2015 (Katolicka Agencja Informacyjna 2022) to 338 in 2022 (Katolicka Agencja Informacyjna 2022). Also, the percentage of students attending school-thought religion declined from 93 percent in 2010 and 70 in 2018 to only 54 in 2021 (Grabowska 2022a). There is also a significant drop in the number of people seeing religiousness as an important value to be taught by the parents: from 29% in 2009 to 16% in 2022 (Bożewicz 2022). Therefore, the above-identified discourses may be used to question and challenge this unwanted positioning of the Catholic Church, thereby supporting social change.

References

Adamczuk, Lucjan. 2001. Kościół w świadomości społecznej Polaków. In Zdaniewicz, Witold (ed.), *Religijność Polaków 1991-1998*, 148-162. Warszawa: Instytut Statystyki Kościoła Katolickiego – Instytut Wydawniczy Pax.

Anthony, Laurence. 2014. *AntConc.* Tokyo: Waseda University. (https://www.laurenceanthony.net/software) (Accessed 2012-03-21.)

Baniak, Józef. 2007. *Desakralizacja kultu religijnego i świąt religijnych w Polsce: Studium socjologiczne.* Kraków: Zakład Wydawniczy Nomos.

Baker, Paul & McEnery, Tony. 2005. A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of language and politics* 4(2). 197-226.

Bednarek, Monika. 2008. Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory* 4(2). 135. doi:10.1515/CLLT.2008.006.

Boguszewski, Rafał. 2009. *Wiara i religijność Polaków dwadzieścia lat po rozpoczęciu przemian ustrojowych: Komunikat z badań*. Warszawa: CBOS. (https://cbos.pl/SPISKOM.POL/2009/K_034_09.PDF) (Accessed 2012-05-14.)

Boguszewski, Rafał. 2016. *1050. rocznica chrztu Polski: Komunikat z badań*. Warszawa: CBOS. (https://www.cbos.pl/SPISKOM.POL/2016/K_056_16.PDF) (Accessed 2022-05-12.)

Boguszewski, Rafał. 2017a. *Przynależność Polaków do ruchów i wspólnot religijnych: Komunikat z badań*. Warszawa: CBOS. (https://www.cbos.pl/SPISKOM.POL/2017/K_084_17.PDF) (Accessed 2022-05-12.)

Boguszewski, Rafał. 2017b. *Zasady moralne a religia: Komunikat z badań*. Warszawa: CBOS. (https://www.cbos.pl/SPISKOM.POL/2017/K_004_17.PDF) (Accessed 2022-05-12.)

Borowik, I. 2002. The Roman Catholic Church in the process of democratic transformation: The case of Poland. *Social Compass* 49(2). 239-252.

Bożewicz, Marta. 2022. *Polacy o wychowaniu dzieci. Komunikat z badań.* Warszawa: CBOS (https://www.cbos.pl/PL/publikacje/raporty_tekst.php?id=6504) (Accessed 2022-10-31.)

CBOS. 2022. Trendy: Praktyki religijne. (https://www.cbos.pl/PL/trendy/trendy.php?trend_parametr=praktyki_religijne) (Accessed 2022-05-12.)

Forchtner, Bernhard & Kølvraa, Christoffer. 2012. Narrating a 'new Europe': From 'bitter past' to self-righteousness? *Discourse & Society* 23(4). 377-400. doi:10.1177/0957926512441108.

Freake, Rachelle & Gentil, Guillaume & Sheyholislami, Jaffer. 2010. A bilingual corpus-assisted discourse study of the construction of nationhood and belonging in Quebec. *Discourse & Society* 22(1). 21-47. doi:10.1177/0957926510382842.

Gabrielatos, Costas & Baker, Paul. 2008. Fleeing, sneaking, flooding. *Journal of English Linguistics* 36(1). 5-38. doi:10.1177/0075424207311247.

Głowacki, Antoni. 2019. *Oceny sytuacji Kościoła katolickiego w Polsce: Komunikat z badań*." Warszawa: CBOS. (https://www.cbos.pl/SPISKOM.POL/2019/K_101_19.PDF) (Accessed 2022-05-12.)

Grabowska, Mirosława. 2013. *Religia i Kościoł w przestrzeni publicznej: Komunikat z badań*. Warszawa: CBOS. (https://www.cbos.pl/SPISKOM.POL/2013/K_170_13.PDF) (Accessed 2022-05-12.)

Grabowska, Mirosława. 2015. *Boskie i cesarskie: O stosunkach między państwem i kościołem(ami): Komunikat z badań.* Warszawa: CBOS. (https://cbos.pl/SPISKOM.POL/2015/K_048_15.PDF) (Accessed 2022-05-12.)

Grabowska, Mirosława. 2022a. *Polski pejzaż religijny – z bliższego planu*. Warszawa: CBOS. (https://www.cbos.pl/SPISKOM.POL/2022/K_091_22.PDF) (Accessed 2022-10-30.)

Grabowska, Mirosława. 2022b. *Postawy wobec obecności religii i Kościoła w przestrzeni publicznej: Komunikat z badań*. Warszawa: CBOS. (https://www.cbos.pl/SPISKOM.POL/2022/K_003_22.PDF) (Accessed 2022-05-12.)

Graff, Agnieszka. 2003. Lost between the waves? The paradoxes of feminist chronology and activism in contemporary Poland. *Journal of International Women's Studies* 4(2). 100-116.

Graff, Agnieszka. 2010. Efekt magmy, czyli o szczególnej roli Kościoła katolickiego w Polsce. In Ostolski, Adam (ed.), *Kościół, państwo i polityka płci*, 10-25. Warszawa: Heinrich Böll Stiftung.

Graliński, Filip & Jassem, Krzysztof & Junczys-Dowmunt, Marcin. 2013. PSI-Toolkit: A natural language processing pipeline. In Przepiórkowski, Adam & Piasecki, Maciej & Jassem, Krzysztof & Fuglewicz, Piotr (eds.), *Computational Linguistics: Applications*, 27-39. (Studies in Computational Intelligence 458). Berlin/Heidelberg: Springer.

GUS. 2011. *Struktura narodowo-etniczna, językowa i wyznaniowa ludności Polski: Narodowy Spis Powszechny Ludności i Mieszkań 2011*.

Heinen, Jacqueline & Portet, Stéphane. 2010. Reproductive rights in Poland: When politicians fear the wrath of the Church. *Third World Quarterly* 31(6). 1007-1021.

ISKK. 2020. *Annuarium Statisticum Ecclesiae in Polonia: Dane za rok 2020*. Warszawa: ISKK. (https://www.iskk.pl/images/stories/Instytut/dokumenty/annuarium-statisticum-ecclesiae-in-polonia-dane-za-rok-2020.pdf) (Accessed 2022-05-12.)

Katolicka Agencja Informacyjna. 2022. *Ilu kandydatów rozpoczęło w tym roku formację seminaryjną?* Warszawa. (https://www.ekai.pl/ilu-kandydatow-rozpoczelo-w-tym-roku-formacje-seminaryjna/) (Accessed 2022-10-30.)

KEP. 2009. *Statut Konferencji Episkopatu Polski*. (https://episkopat.pl/statut-konferencji-episkopatu-polski/) (Accessed 2022-05-12.)

Kominek, Andrzej. 2003. *Punkt zborny: Obraz Kościoła katolickiego w Polsce w tekstach publicznych z lat 1970-1989.* Kielce: Akademia Świętokrzyska im. Jana Kochanowskiego.

Makuchowska, Marzena. 2011. *Od wrogów do braci: Posoborowe zmiany w dyskursie Kościoła katolickiego.* Opole: Wydawnictwo Uniwersytetu Opolskiego; Uniwersytet Opolski.

Mandes, Sławomir. 2020. Clerical sexual abuse in an illiberal state: The case of Poland. *Journal of Church and State* 62(1). 110-134. doi:10.1093/jcs/csz089.

Mariański, Janusz. 1993. *Religia i Kościół w społeczeństwie pluralistycznym: Polska lat dziewięćdziesiątych.* Lublin: Katolicki Uniwersytet Lubelski.

Mariański, Janusz. 2011. *Katolicyzm polski – ciągłość i zmiana: Studium socjologiczne.* Kraków: Wydawnictwo WAM.

Mishtal, Joanna & Dannefer, Rachel. 2010. Reconciling religious identity and reproductive practices: The Church and contraception in Poland. *The European Journal of Contraception & Reproductive Health Care* 15(4). 232-242. doi:10.3109/13625187.2010.498595.

Mishtal, Joanna Z. 2009. Matters of "conscience": The politics of reproductive healthcare in Poland. *Medical Anthropology Quarterly* 23(2). 161-183.

Owczarzak, Jill. 2009. Defining democracy and the terms of engagement with the postsocialist Polish State: Insights from HIV/AIDS. *East European Politics & Societies* 23(3). 421-445.

Pakuła, Łukasz P. & Pawelczyk, Joanna & Sunderland, Jane. 2015. *Gender and sexuality in English language education: Focus on Poland.* London: British Council.

Salama, Amir H.Y. 2011. Ideological collocation and the recontexualization of Wahhabi-Saudi Islam post-9/11: A synergy of corpus linguistics and critical discourse analysis. *Discourse & Society* 22(3). 315-342. doi:10.1177/0957926510395445.

Scovil, Jonathan. 2021. *Stosunek Polaków do osób homoseksualnych: Komunikat z badań*. Warszawa: CBOS. (https://www.cbos.pl/SPISKOM.POL/2021/K_121_21.PDF)

Skowronek, Katarzyna. 2006. *Między sacrum a profanum: Studium językoznawcze listów pasterskich Konferencji Episkopatu Polski (1945-2005).* Kraków: Wydawnictwo Lexis.

Stubbs, Michael. 2005. *Words and phrases: Corpus studies of lexical semantics.* Malden–Oxford–Carlton: Blackwell.

Zaręba, Sławomir H. 2010. Instytucjonalny wymiar religijności. In Zaręba, Sławomir H. (ed.), *Rodzina, religia, społeczeństwo: Polacy 2009 w diagnozie socjologicznej*, 75-138. Warszawa: Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie.

Zielinska, Eleonora. 2000. Between ideology, politics, and common sense: The discourse of reproductive rights in Poland." In Gal, Susan & Kligman, Gail (eds.), *Reproducing gender: Politics, publics, and everyday life after socialism*, 23-57. Princeton: Princeton University Press

On the prospects of a logical-semantic approach to creating a code for artificial thinking

Karen Kocharyan[1] & Włodzimierz Lapis[2]

[1] [No affiliation]
svbraun@rambler.ru | ORCID: 0000-0001-9114-5608
[2] Adam Mickiewicz University in Poznań
lapisw@amu.edu.pl | ORCID: 0000-0002-6360-4018

Abstract: This article substantiates the practical possibility of modeling the information component of thinking using modern computer technologies, providing programmers with a code diagram of general artificial thinking for familiarization. The rationale for the approach is based on the ideas of the theory of knowledge, philosophy and methodology of language. Attention is paid to the discussion of some provisions of the criticism of artificial systems of thought.

Keywords: artificial thinking, logical-semantic approach, information modeling, philosophy of language, data processing, reality visualization, cognitive linguistics, generative linguistics, linguistic competence, informational modeling

In this article, we will attempt to discuss the prerequisites for possible explanations of artificial thinking mechanism from the standpoint of its connection with language and reality, in order to determine the logical-semantic approach to describing the architecture of the information model of thinking, justified by the modern level of linguo-philosophical knowledge.

These explanations and descriptions require some generalization of well-known linguo-philosophical paradigms and are aimed at the practical implementation of the proposed architecture in a real software product, and at a subsequent verification of its functionality.

We saw the task of linguistic knowledge in the era of the information revolution to provide programmers developing strong artificial intelligence technologies with the tools of linguistic philosophy. The solution to this problem is to substantiate the logical-semantic approach to the representation of the artificial thinking code diagram, the construction of which follows from the main well-known areas of linguistic theory and linguistic philo-

sophy, and to propose a namespace for variables, functions and modules that determine the content of the code both from theoretical positions and from positions of practical coding.

However, first, it is necessary to explain the basis of our approach to creating a model of artificial thinking, since there is a reasonable criticism of the possibility of the appearance of a thinking computer by linguists and philosophers we respect.

The sharpest and most just argument in this criticism is the description of the "Chinese Room" by J. Searle (1993: 3-66). From this follows the conclusion that it is impossible to get an idea of reality only by processing the data of the sign system. And indeed it is. However, we must remember that the year of the first publication of the "Chinese Room" is 1980, and the speed of data processing at that time did not allow digital imaging to be included in the fast processing process. Now, more than 40 years later, the speed of data processing allows processing the simultaneously presented sign of the language and the data of the surrounding visual world, which, in our opinion, is a key requirement for reflecting discourse. Visualization and/or replacing/supplementing it with the data of the surrounding reality solves the main task, which is predetermined by the theme of the "Chinese Room" – to connect the processing of the data of the sign and his designate. It is the process of generating the unity of the sign and his designat, which solves the problem of the "Chinese room", and, then, by this article we intend to initiate the usage of the information component of this unity in artificial thinking. Let us recall, in this regard, the key conclusions of Saussure's semiotics (1933: 55-197): language can be represented as a system of relational dependencies, and language is a system of conventional signs. The data processing of the sign system in the "Chinese Room", Searle spoke about, raises the problem of determining only relational dependencies and does not touch upon the issue of conventionality, since the conventional nature of the sign of a language has as its subject the unity of the signifier and the signified, while the relational nature of language has only a sign system as its subject.

Linking the processing of language and the visual part of reality into a single whole by means of simultaneously processed speech and visual signal, however, we immediately face the following justified criticism of any attempts to obtain artificial intelligence. These are the factor of "transcendence", in terms of Kant (1964: 585-695), and the factor of participation of "non-objectifying", in terms of Husserl (Prechtel 1999: 23-29), not substantiated by the reality of mental acts in human thinking, those are the subject of constant linguists disputes about the object of linguistic knowledge, rooted the unknowable part of brain activity. Brain processes are not directly observable, so there is no point in guessing about what is, in principle, inaccessible - this is the main thesis of antimentalist-minded descriptivists. Bloomfield believed that "the descriptive order of grammatical features is generated by the method of describing linguistic forms" (Bloomfield 1933: 213). To interpret the external behavior of an individual in terms of brain/psychic entities is to fall into a logically vicious circle: to explain the visible through an unknown cause. Authoritative representatives of analytic philosophy echo this argument: We "should never be concerned with reducing the one to the other, nor explaining the one in terms of the other. Philosophy is indeed purely descriptive" (Wittgenstein 1964: 18).

However, the mentalists were not powerless in solving this problem. It is known that the relatively recent cognitive trend in linguistics raised questions about what are the

general integral foundations of the processes of cognition and what manifestations of the unknowable part of thinking can be studied. Methods for detecting such grounds and fixing such manifestations are information-theoretic modeling. Moreover, the leading methodology of the cognitive sciences is the informational approach, which considers reality from the point of view of the information processes that take place in it, and "a person is understood as a thinking system in the world of information" (Baksanskii 2005: 82-83). Generative linguistics, which is a special object of our attention when constructing a code of artificial thinking, is also in dire need of mental reality, since it relies on the category of competence, in which mental states perform a causal or generative function in relation to observed communicative processes, and events, which include mental constructs can function as links in a causal chain. One of the leading theorists of rationalism, J. Katz, explains the essence of causal mentalism, emphasizing that "if the logical consequences of the model are consistent with the observed behavior of the system and would not do this without the hypothesis put forward, then the scientist has the right to say that the hypothesis explains the behavior of the system in terms of the observed, but a causally effective component" (Katz 1965: 128). The linguist builds a theory on the basis of a hypothetical conclusion about the characteristics of the mechanism underlying communication. Believing that the observed events of linguistic communication are a consequence of this mechanism, he proposes a theory about the structure of this mechanism and a causal chain connecting it with observed events to explain how these internal causes give rise to linguistic communication as their effect. The ability to model the informational part of thinking with modern computer tools brings to the fore not only information-theoretical modeling, which currently has a central place in the methodology of cognitive linguistics, but also practical information modeling, which expands the paradigm of cognitive knowledge: we get a methodological basis for explaining subjectively-oriented (anthropological factor) of ordinary cognition through agent-oriented (computer factor) cognition, implemented in a specific algorithmic code.

So, we are not going to model consciousness in order to create a mechanism of artificial thinking: its processes are not known. We are going to model thinking, and do it from the perspective of the information component of thinking. We must make several important explanations in this regard.

The first explanation is about the linguistic reflection of the non-objectifying component of human thinking by the code of artificial thinking. This component is assessments, desires, emotions, that is, intentional non-objectifying acts manifested in the signs of the language. At the same time, desires, emotions, assessments, as a meaningful result of these acts, are represented by a significant part of the human lexicon, that is, they are designated as any real object is designated, and their designation has the same conventional nature, any language sign has. A single apparatus of artificial thinking should not be endowed with complete information of human sensory perception or a mechanism of empathy. Not all of us are astronauts, but all of us, from a certain age, can use the word "weightlessness", operating with it with the proper level of relevance. The same can be said about the word "drugs" and many other words. Obviously, in the minds of not all people, the complete conceptual composition of the language is represented using sensory data at a non-objectifying level. But, however, all the richness of the native language

is available to the thinking of the native speaker of this language. How exactly does this happen? We conclude that it is possible for artificial thinking to objectify, for example, the concept of "pain" with discursive data of the corresponding behavior of its main user, as we do when linking the concept of "weightlessness" and the broadcast from the orbital station that we once saw.

The second explanation is about the role of the material basis for the production of the thought process. In our project, we proceed from the conclusions of the relativistic direction of linguistics and semiotics about the nature of the sign, the signified and their unity. The key role in the functioning of a linguistic sign is played not by its material basis, but by its distinctive meaning, manifested in its relationship with another sign and signified meaning. Illustrating the irrelevance of substance for the functioning of language as a structure, Saussure says that the replacement of wooden pieces by ivory pieces (substantial qualities) in chess is indifferent to the system. But, if you increase or decrease the number of figures (relational qualities), then this "will deeply affect the 'grammar' of the game". (Saussure 1933: 45).

The following explanation concerns the "question of truth", in terms of Tarski (1972: 136-145), the solution of which is important when operating with a conventionally justified sign of the language in order to implement the very mechanism of mental reaction to the current reality. "Existent substance and nomination", in the terms of Arutyunova (1976: 205-283), generated on the concrete-referential basis of the visual series, manifest "apodictic evidence" in terms of Husserl (Prechtel 1999: 37-40) of existence as an independent value/significance, in turn generate relation to the truth of subsequent processes of predication and identification. Both of the last mentioned processes are included in the function of generating data of a universal, and then an abstract reference, which is not directly related to the current composition of the incoming data. The concrete-referential basis of the processes of predication and identification in the discourse of visual content lays the foundation for the truth of thinking in general, which includes abstract concepts – ideas, that are not directly correlated with the surrounding visual reality. Thinking in general, in the context of the informational approach, can be understood as augmented reality, where the concrete-referential content of predicative constructions tied to apodictic evidence is filled with universal and abstract concepts used by thinking. The discourse itself at a certain stage in the development of thinking appears as an augmented reality in which the plan of content does not coincide with the plan of expression. However, this does not destroy the doxal nature of thinking, which permanently evaluates reality from the standpoint of truth.

The creation of a descriptive and explanatory basis for the general architecture of artificial thinking is aimed at revealing the content of the meaning formation process, what Chomsky (1968: 5) explains as "innate knowledge" in his theory of competence. We believe that the content of the process of meaning formation is represented by the predicative-discursive structure of data segmentation. For all data coming from the visual and speech stream, the code finds a place in the predicative part (Fig. 1), and in the discursive part of the structure (Fig. 2). This is the top level of data processing. It is preceded by the segmentation of visual stream data in the "Representamen" structure component in the phenomenological sense of Pierce (1953: 212), as that "information

about the essence that the phenomenon carries in itself" and speech stream data in the "Glossema" structure component in terms of Hjelmslev (1960: 66-337), as "the shortest unit of linguistic meaning" (Fig. 3). The work of these structures is ensured by the basic principle of the code functioning: the principle of cyclical movement of information. The principle of cyclic movement provides a constant flow of interpretive knowledge into the interpreted and vice versa. At the stage of movement from the visual-speech flow to the predicative-discursive structure, the knowledge previously accumulated in the areas of "Representamen" and "Glossems" is interpretive, and the new information at the stage of its segmentation in the current predicative-discursive data structure is interpretable knowledge. The signal of relevant segmentation of data in this structure is the discovery of the previously known environment of these data in the paradigmatic hierarchy of reality, built on the theory of "hard designators" Kripke (1986: 194-242) and in syntagmatic sequences. Otherwise, the code is forced to form a new knowledge paradigm, in which the new information will be relevant and will receive a new truth evaluation matrix. After the relevance has been verified, the new information has fulfilled its role as a signified (interpreted knowledge) and is ready for the role of a signifier (interpretive knowledge) before a new cycle. Thus, the generative essence of language competence described by Chomsky is implemented in the code.
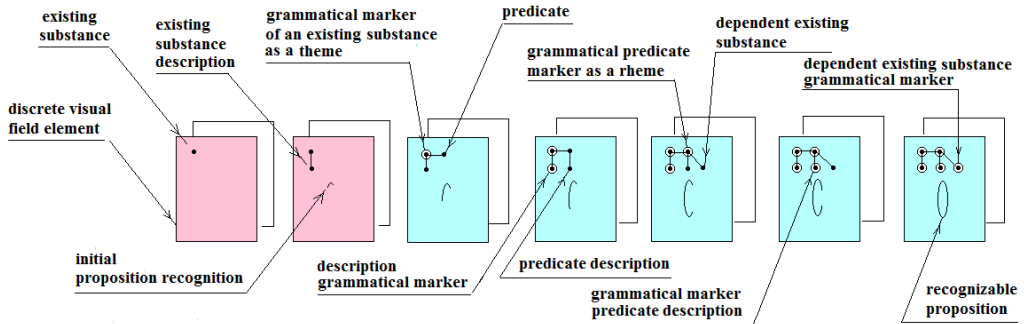
In order to implement the proposed model, it is necessary to further develop specific technological operations. Thus, an essential technological aspect of cyclic information processing that requires further development is the description of the data that form the beginning of the cycle and the level of transformation of this data at the end of each cycle. For practical implementation in the code of the matrix of the truth, it is also necessary to further study the issue of the information technical division principles in accordance with the specified markers of doxality "fantasy (absence of doxality)–denial–doubt-assumption–belief–conviction–truth". If the first and last elements of the given scale can rely on a strict algorithmic procedure for the correspondence of the assessed information to the data that have received the status of interpretive, then the technological basis for determining the correspondence of the assessed information to the intermediate markers requires additional development.

The code diagram shows the process of initiating linguistic communication. Based on the current discourse, the code forms the target discourse, mobilizing, on the one hand, the data of the possible continuation of the current discourse with the data of discursive sequences and the paradigmatic hierarchy of the truth picture of the world, and, on the other hand, the data of discursive sequences of the modal-evaluative content of the main user and the paradigmatic hierarchy of his priorities. The transition formed by the memory data in the form of a discursive sequence from the current discourse to the target one forms the basis of the signified in the mental reaction.The current situation is only a new way of ordering the old, already existing material. This amazing thesis was once expressed by Marcel Proust, who said that when, for example, one reads his book, then "with the help of his book one reads oneself" (Proust 1927: 201).

The code diagram of artificial thinking, giving an answer to the question of how exactly a thought arises and develops in discourse, defines the basis of a software product with its namespace, the terminological essence of which we strive to convey to coding programmers.

Figure 1. The predicative structure formation

Figure 2. Discourse formation process

At this stage, the proposed model is theoretical in nature. For its implementation, it is necessary to confirm it with experimental results and a prototype, which is being worked on.

Artificial thinking code diagram ˙
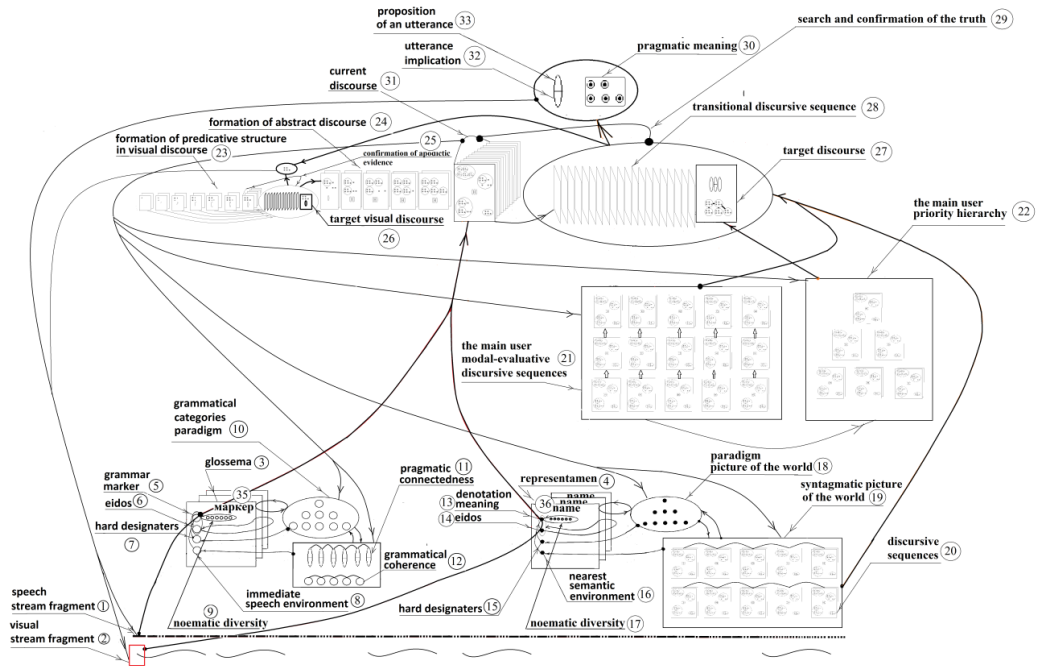
Figure 3. Artificial thinking code diagram

We believe that this diagram approximates our understanding of metalanguage, that is, that element of Kant's "a priori", Chomsky's ideas about the "linguistic innateness" of a person, which allow people to acquire any language as the basis of empirical knowledge. Taking as a basis Saussure's brilliant conclusion: "the whole linguistic mechanism revolves around identities and differences" (Saussure 1933: 109), we show how visually completely dissimilar things (visual data) acquire something in common, equivalent, which absolutely equalizes them, turns them into identity (certain component of predicative structures and discourse). In our view, this process realizes the emergence of linguistic value in the understanding of Saussure. The technology of building a presupposition of a holistic discourse is associated with the emergence of predicative structures, the components of which can themselves be expressed by predicative structures that do not have a formal visual connection with the current discourse, maintaining abstract concepts. We proceed from the fact that the holistic discourse of the ideal speaker/listener is abstract. The very formation of the presupposition of abstract discourse by predicative structures of visual

discourse, which previously realized the grammatical structure of the language, is based on the Sapir-Whorf hypothesis of recognizing the relationship between thinking and grammatical categories, on Whorf's statement that "the fundamental system of language (in other words, grammar) is not just a means of expression ideas, but a means of forming ideas" (Whorf 1956: 212-213), that is, a means of abstract thinking.

This article does not consider the ethical side of the model implementation. We consider the accumulation of data revealing the subtle psychological characteristics of each individual user to be a significant ethical problem associated with the creation of artificial thinking systems using the proposed model. Since it is in accordance with the identified aspects of the discursive behavior of an individual user that the proposed model can form the composition of interpretive data of the vocabulary of intentional content. Taking into account potential biases and social consequences of this factor requires separate coverage.

References

Arutjunova, N. 1976. *Predloženie i ego smysl: Logiko-semantičeskie problemy.* Moskva: Nauka.

Baksanskij, O. 2005. Nejrolingvističeskoe programmirovanie kak praktičeskaja oblast' kognitivnyh nauk. *Voprosy filosofii* 1. 82-100.

Bloomfield, L. 1933. *Language*. New York: H. Holt.

Chomsky, N. 1968. *Language and mind*. New York: Harcourt, Brace&World.

Hjelmslev, L. 1960. Prolegomeny k teorii jazyka. *Novoe w lingvistike*. Vyp. 1. 66-337. Moskva: Inostrannaja Literatura.

Kripke, S. A. 1986. Zagadka kontekstov mnenija. *Novoe v zarubežnoj lingvistike* 18. 194-242.

Kant, I. 1964. *Kritika* čistogo razuma. Sočinenija v 6 tomah, vol. 3. Moskva: Mysl'.

Katz, J. J. 1965. *Mentelism in linguistics*. Language 4. 124-137.

Pierce, Ch. S. 1953. *Letter of 23 December 1908.* In *Letters to Lady Wilby*, edited by I. C. Lieb, p. 212. New Haven: Whitlock.

Prehtl', P. 1999. *Vvedenie v fenomenologiju Gusserlja.* Tomsk: Vodolej.

Proust, M. 1927. *A la recherche du temps perdu*. Paris.

Serl', D. 1993. *Soznanie, mozg i nauka*. Put' 4. 3-66.

Sossjur, F. De. 1933. *Zametki po obščej lingvistike*. Moskva: Socekgizm.

Tarskij, A. 1972. Istina i dokazatel'stvo. *Voprosy filosofii* 8. 136-145.

Whorf, B. 1941-1956. *Language, thought and reality.* Cambridge: MIT Press.

Wittgenstein, L. 1964. *The blue and brown books: Preliminary studies for the 'Philosophical investigations'.* Oxford: Blackwell.

Bartholomae's law revisited and remodelled

Ondřej Šefčík

Masaryk University, Brno

sefcik@phil.muni.cz | ORCID: 0000-0001-6368-9421

Abstract: The present paper deals with Bartholomae's Law in Indo-Iranian languages. After a new examination of the data, it concludes that in the case of Bartholomae's Law, there is no assumed ´transfer of aspiration' or any other similar process but that the attested plosives are the result of a process of plosivization of the original Indo-Iranian voiced fricatives Δ (< IE *D^h). Thus, the process described by Bartholomae is not an exception in developing of plosive clusters but a regular process affecting the original fricatives, plosives in the outcome clusters being of secondary origin.

Keywords: Bartholomae's Law, voiced aspirates, voiced spirants; Indo-European; Indo-Iranian

1. On Bartholomae's Law in general

Observing the development of the Indo-European clusters formed by two obstruents into Indo-Iranian,[1] we see that the typical assimilation process affecting these clusters is **regressive** in its orientation, the left[2] standing obstruent being assimilated both in voice and aspiration to the right standing obstruent. However, the clusters formed by the left IE voiced aspirate *D^h and a voiceless plosive on the right are remarkable exceptions since they are subject to the assimilation process of a **progressive** orientation where the right unaspirated voiceless plosive becomes voiced in both Indo-Iranian branches and also aspirated in Indic, the process described by **Bartholomae's Law**.[3] In Iranian, even the

---

[1] Though some authors consider the Law to be operating already in IE (cf. Kuryłowicz 1935: 50-51; Lubotsky 2018: 1879), we support the idea that the Law is exclusively Indo-Iranian (e.g., Szemerényi 1990: 107; Hoffmann & Forssmann 1996: 95-96). Noteworthy is that the possible validity of Bartholomae's Law for Germanic was examined in recent years, especially by Görtzen (1998: 444-448) and Hill (2003: 218-220).

[2] The terms *left* and *right* are here used purely conventionally to describe both an earlier (left) and later (right) pronounced element.

[3] Named after Christian Bartholomae, who described the mechanism first (for details and history of the law, see below).

right sibilant become voiced as well (the question of original aspiration of both types of right obstruents in Common Indo-Iranian will be also discussed).

> **NB:** For simplicity, we use the traditional IE triad *T – *D – *D^h (used since Lehmann 1952) as purely algebraic symbols, without any attempt to attach any phonetic values to them, as we are willing to refuse to use any of many alternative models (ʻglottalicʼ, ʻejectiveʼ etc.). What is remarkable both *T and *D behave in the examined clusters in the same way, with the same outcomes, unlike *D^h; hence, the unique symbol T/D will be used in formulae.

Our analysis will focus only on clusters formed by the right-standing *t- and *$s/š$-[4] (ʻBartholomaean clustersʼ) since these are generally the most common both in derivative and inflective morphs; the examples we will use are those of the verbal derivation and inflexion since securely attested and synchronically productive.

Analogical clusters formed by a right standing *d^h-, though not strictly Bartholomaean,[5] will be examined as well since they represent the functional background for the above-mentioned Bartholomaean clusters; hence, the clusters we will examine are:

i. formed by an IE left obstruent (*T/D,[6] *D^h);
ii. followed by a right obstruent: *t-, *d^h-, *$s/š$-).

There are six possible combinations, including either Bartholomaean or non-Bartholomaean clusters: *T/D + *$t/d^h/s$ and *D^h + *$t/d^h/s$.

2. Bartholomae's Law in Indo-Iranian

As we wrote above, Bartholomae's law is well attested as a synchronic process in both Old Indic and Iranian (Avestan). However, the two branches differ in the number of clusters affected by the process and the outcomes.

> **NB:** Old Persian does not have any attested ʻBartholomaean outcomesʼ, the cluster *$D^h t$ has only a non-Bartholomaean regressive outcome TT (cf. OP ppp. *basta*- derived from the √*band*- ʻbindʼ cf. LIV^2 75-76; Cheung 2007: 4-6; Brust 2018: 265; ppp. *gasta*- from √*gant*- ʻstinkʼ; cf. Cheung 2007: 103-104; Brust 2018: 177-178), the cluster *$D^h S$ has no known outcomes either Bartholomaean or not. In Old Persian attested clusters *T/Dd^h and *$D^h d^h$ are naturally not subject to Bartholomae's Law, resulting from the (expected) regular regressive process (cf. OP adv. *azdā* if from *√*ad^h*-; cf. LIV^2 222; Lipp 2009b: 87; Cheung 2007: 153; Brust 2018: 132-133).

The differences between the two branches are as follows:

---

[4] Both sibilants are just positional (according to the *ruki*-rule/*Pedersen's Law*) variants of a single sibilant.

[5] Non-Bartholomaean clusters with *d^h- are subject to regressive assimilation but share outcomes with the Bartholomaean *t- clusters, as shown below.

[6] OIA voiceless aspirates never enter the clusters in the left position, being always separated from the right obstruents by an inserted vowel (cf. Šefčík 2012).

In **Indic**, only clusters resulting from *$D^h t$ were affected (marked **bold**). The clusters *$D^h s$ have, unlike in Avestan, newly created regressive non-Bartholomaean outcomes (marked *cursive*):[7]

Table 1

OIA	*t-	*s-	*d^h-
-*T/D	Tt	Ts	Dd^h
-*D^h	**Dd^h**	*Ts*	Dd^h

On the contrary, Bartholomae's Law affects two types of clusters (marked **bold**) in **Avestan**: *$D^h t$ and *$D^h s$ (see the table with the analogous non-Bartholomaean clusters):[8]

Table 2

Av.	t-	s-	d^h-
-*T/D	Θt	(Θ)s	Dd
-*D^h	**Δd**	**Δz**	Dd

However, Avestan also exhibits a large number of analogous regressive outcomes besides the regular Bartholomaen (*Tt* and *Ts* instead of *Δd* and *Δz*); parallelism, on the contrary, is entirely unknown in Indic, which always has only one type of output.[9]

Another essential difference between the two branches is the one affecting the location of the left plosive: in Iranian, the left peripheral (velar or labial) plosive is fricativized before *t- but in Old Indic, the plosive is preserved. In Iranian, the left central (dental or originally palatovelar) plosive is sibilantized; in Old Indic, only palatovelars are realized as sibilants (as in other *satəm*-languages).

For the reconstructed Common Indo-Iranian state, we assume that:

i. there was an original IE triad *T, *D, *D^h,[10] attested still later as a triad in Indic and with its third member neutralized on D in Iranian; however, Bartholomae's Law shows that Indo-Iranian had this triad still preserved (and probably accompanied by a newly created *T^h) and the fusion of both voiced classes is a secondary and later feature of Iranian; The old Indo-European triad *T, *D, *D^h was still retained in Indo-Iranian regarding the number of opposition members, plus the addition of the newly formed Indo-Iranian *T^h. However, for IE *D^h, we assume a different phonetic value in Indo-Iranian, namely *$Δ$, as shown below. Later, in Iranian, IE *D and *D^h merged into a single syllable (Ir. *D), and *T^h became a spirant (Ir. *$Θ$). In Indo-Iranian, on the other hand, the phonemic status of IE

---

[7] For the given clusters and their OIA outcomes, see Appendix I.

[8] For the given clusters and their Avestan outcomes, see Appendix II.

[9] Here, we are setting aside unique outcomes of the root √*dhā*- 'put' as: pr. imp. *dhattá*(*na*), impf. *ádhattam*, etc., areexceptions in otherwise regular OIA developments.

[10] These symbols are used purely algebraically and used here only to distinguish between three modal classes.

$*D^h$ was reshaped to complement the new Indo-Iranian $*T^h$ as a new fourth member in a system of oppositions based on voicing and aspiration.

ii. Iranian $\varDelta z$ from IE $*D^h s$ represents an older state of arts; Indic outcome Ts for the same input is a later levelling according to the regressive neutralization (similar neutralization of Bartholomaean clusters appears in Iranian as well as variants to the Bartholomaean development); hence both clusters $*D^h t$ and $*D^h s$ were subject to Bartholomae's Law.

Table 3

InIr	t-	s-	d^h-
-T/D	Tt	Ts	Dd^h
-D^h	**Dd^h**	**Dz**	Dd^h

NB: For simplicity, we leave the question of the fricativization of the left plosives (of the dental and old palatovelars series) aside at this moment, just stating that:

i. the development of dentals was already the Late Indo-European and Indic state of arts (with the left dental plosive) is a secondary archaization, not an older state of arts;

ii. the development of palatovelars was already common for all *satəm*-languages.

3. Previously used models of Bartholomae's Law

As noted above, the progressive direction of Bartholomae's Law goes contrary to the prevailing tendency of Indo-European (and Aryan as well) clusters to be assimilated in a regressive manner (note that the outcomes of the Law are often later analogically levelled on 'regular' regressive outcomes – regularly in Indic for clusters $*D^h S$ to point out a very remarkable example, however very often analogical regressive outcomes are present in Avestan beside regular progressive outcomes, according to Bartholomae's Law).[11]

The law was first described by Bartholomae (1882), who later returned to it many times (Bartholomae 1883: 48; Bartholomae 1885: 206; Bartholomae 1895-1901: 21-23), and the law was repeatedly investigated, revalued and remodelled since. For further references to older literature, see especially Collinge 1985: 7-11 and Mayrhofer 1986: 115-
-118; Szemerényi 1990: 106-109; Mayrhofer 2004: 46).

Bartholomae stated that the result of the concatenation of any left voiced aspirated obstruent and any right voiceless unaspirated obstruent would be a cluster of a voiced and a voiced aspirated obstruent (including aspirated sibilant), schematically: $*D^h T >$ $*DD^h$, $*D^h S > *DZ^h$. His description only observes the inputs and outputs; there is no detailed process model.

Other authors proposed different models:

i. Anderson (1970: 388) assumes two phases: in the first, a cluster of two voiced aspirated plosives are created, and the second phase is a deaspiration of the first plosive ($D^h T > D^h D^h > DD^h$);

---

[11] Cf. IE *ts* > OIA *ts*, Av. *θs*; IE *dʰs* > OIA *ts*, Av. *θz*; IE *Ḱs* > OIA *kṣ*, Av. *θš*; IE *ǵʰs* > OIA *kṣ*, Av. *θž*.

ii. the deaspiration as a second step of the whole process is accepted by Schindler (1976), who assumes that an obstruent becomes aspirated (and inherently voiced) after a voiced aspirate, though other authors assume that Indian aspirates are of biphonematic nature (cf. Ejerhed 1981);

iii. Sag (1974: 593) states there is a paradox: Bartholomae's law (and subsequent deaspiration) should, according to him, precede Grassmann's Law in case of *bhot-sya-* but Grassman's Law should precede Bartholomae's Law in case of *buddha-* (also cf. Sag 1976);

iv. Mey (1972) forms a complex of processes, where a deaspiration with a subsequent devoicing before an obstruent goes through a series of shifts of inter-exclusive operations;

v. D. G. Miller (1977) assumes the influence of the root structure on the process, considering the voicing process as a primary trajectory, followed by the aspiration as a later process;

vi. Lombardi (1991: 140) tries to explain the unexpected voicing of the right obstruent and the transfer of aspiration to it as 'spreading of the entire Laryngeal node', i.e., by aspiration of the whole cluster;

vii. Kobayashi (2004: 117-125) speaks about the 'aspiration throw-back', using the instrumentality of the Optimality theory and following the morphemic structure of clusters;

viii. De Angelis (2006) recalls older Schindler's opinion (Schindler 1976: 629), assuming that the input cluster $*D^hT$ aims to preserve the aspiration by shifting it on the left plosive. Again, this model is based only on Indic data and does not explain why the assumed shift of aspiration on the $*T$ is accompanied by the shift of voice, especially since Indic has T^h, the possible cluster hence could avoid, if there is any need for the 'shift of aspiration', the basic least marked form is †TT^h, not attested DD^h.[12]

Previous models often focus only on OIA data, and thus only on clusters $*D^hT$, and therefore usually omit the parallel cluster $*D^hS$ (since its outcome in the Indic is 'regular regressive' *TS*, unlike in the Iranian). Similarly, they usually ignore parallel clusters $*TD^h$ and $*D^hD^h$, both sharing the same outcome with $*D^hT$-clusters (i.e., Indic DD^h, Iranian *DD*), which makes them a functional background for Bartholomae's Law (the same outcome is the result of two processes in opposite directions).

As we can see, the typically used model generally assumes a shift of aspiration from the left plosive to the right plosive, accompanied, if considered at all, by various more or less complicated processes and a shift of voice. Such models are implausible, even twice: aspiration and voicedness are **marked** values of obstruents and unaspirated and voiceless are, on the contrary, unmarked values, and the typical assimilation of these values proceeds in the opposite direction, i.e. from marked to non-marked and in a regressive direction.

---

[12] Again, cf. Šefčík (2012) shows that OIA $*T^h$ is never synchronically neutralized or created, unlike other modal classes of plosives.

4. The proposed alternative model

The reconstructed phonetic value of IE and/or Indo-Iranian $*D^h$ needs to be re-evaluated and clarified.[13] In the following lines, we will use IE $*D^h$ as a purely algebraic sign without any comments on its properties in Indo-European. The traditionally reconstructed IE 'voiced aspirates' are realized in Indic as breathy voiced plosives,[14] in Iranian, they are merged with plain voiced plosives (similarly as in Celtic and Balto-Slavic). However, the voicedness is the essential property of Indo-Iranian $*D^h$ (as we can assume for Celtic, Balto-Slavic and even Germanic and Armenian but not for Proto-Greek or Italic).

We assume that IE $*D^h$ was realized in Common Indo-Iranian as a voiced spirant $*\varDelta$, not as an aspirated voiced stop. This idea was suggested already by von Brücke (1856: 59-60), later also by Walde (1887: 466), Prokosch (1918-1919; Prokosch 1939: 39-41) and Hammerich (1967: 839-849), although not specifically for the Bartholomaean clusters but in general for IE $*D^h$ in any position.[15]

Our proposed model for the development of the IE $*D^h t / *D^h s \| *TD^h / *D^h D^h$ clusters (the first pair is Bartholomaean, and the second is parallel to it) follows the trajectory of two steps: the early spirantization of the IE $*D^h$ in Indo-Iranian on $*\varDelta$ in all positions and the subsequent processes of further assimilations and lenitions in clusters. To the development of individual clusters:

1. We assume the following trajectories for $*D^h T$ clusters with the following steps:
    a. the left IE plosive (= $*D^h$) being an Aryan voiced spirant ($\varDelta$) caused the following t to become also a voiced spirant;
    b. in the second phase, both spirants became a subject of fortition to plosives; the left spirant became a voiced plosive, the right spirant changed into a voiced aspirate ($\varDelta\varDelta > DD^h$)[16] in Indic,[17] both plosives are realized as unaspirated plosives in Iranian:

   i. IE $*D^h+T > \varDelta+T > \varDelta\varDelta > DD^h$ (Indic)
  ii. IE $*D^h+T > \varDelta+T > \varDelta\varDelta > DD$ (Iranian)

---

[13] For the phonetic value of IE $*D^h$ cf. Kümmel (2012: 293), who successfully argued against 'voiced aspirated' value. From the phonemic point of view, the triad $T – D – D^h$, though many times attacked as 'typologically improbable' (cf. Jakobson 1958: 22-23 and Hopper 1973: 141), is well attested in Madurese, Kelabit and probably also in Bintulu (cf. Blust 2009: 174-175, 182).

[14] The phonetic value of Indo-Aryan D^h is "breathy voiced", but phonologically, the position of D^h is clear: it is in the proportional opposition to T^h as T is to D (i.e., the opposition of voice), and as much in the proportional opposition to D as T is to T^h (i.e., the opposition of aspiration). Moreover, breathy and aspirated stops are indeed acoustically similar in that there is a delay in the onset of full voicing. Thus the phonetic proximity between breathy phonation and aspiration is apparent.

[15] Brixhe (1997) assumes that Paleo-Balkan languages (including Proto-Greek!) underwent the change IE $*D^h > *\varDelta$.

[16] We assume, similarly to D.G. Miller (1977), that voice was a primary quality, not aspiration.

[17] Either directly due to the same process or later, Walde (1897) assumes aspiration as a later feature for OIA and Greek aspirates.

2. *Simili modo*, the trajectories for the development of the IE **$*D^hS$** (InIr. *$*\Delta S$) clusters are modelled with steps:
   a. the left voiced spirant caused the left sibilant also to become voiced;
   b. the intermediate ΔZ cluster became a plosive in the left part of it in Iranian; *TS* replaces the expected †*DZ* due to analogy with regressive clusters in Indic, the Iranian state is assumed to be archaic, hence:

i. IE $*D^h+S > \Delta+S > \Delta Z (\rightarrow TS)$ (Indic)
ii. IE $*D^h+S > \Delta+S > \Delta Z > DZ$ (Iranian)

NB: The spirantization model of Bartholomae's Law has one prominent advantage concerning the development of $*D^hS$ clusters: within the spirantization model, there is no need to introduce the 'exotic' voiced aspirated sibilants (Z^h) at all.

The development of both analogical clusters can be modelled in the following way:
3. For the development of IE **$*T/D+D^h$** cluster, we simply assume that:
   a. the left plosive took the voice and is spirantized due to the right Δ (< IE $*D^h$);
   b. both spirants were subject to a fortition both in Indic and Iranian:

i. IE $*T/D+D^h > T+\Delta > \Delta\Delta > DD^h$ (Indic)
ii. IE $*T/D+D^h > T+\Delta > \Delta\Delta > DD$ (Iranian)

4. The development of **$*D^hD^h$** clusters we model as follows:
   a. an Aryan left voiced spirant (= IE $*D^h$) forms a homorganic cluster with a right voiced spirant;
   b. in the second phase, both spirants became a subject of fortition into plosives; a left spirant became a voiced plosive, the right spirant changed into a voiced aspirate ($\Delta\Delta > DD^h$) in Indic, the right plosive is unaspirated plosive in Iranian:

i. IE $*D^h+D^h > \Delta+\Delta > DD^h$ (Indic)
ii. IE $*D^h+D^h > \Delta+\Delta > DD$ (Iranian)

The fortition of Indo-Iranian voiced spirants on plosives is significantly rarer than lenition from the typological point of view but not impossible. Here, we have to mention the fortition of loanwords in Goidelic. Even in Indic, we have modern synchronic alternations of $*\Delta$ and $*D^h$ in Gujarati: intervocalic /b^h, g^h, d^h) have variants [β, γ, δ] (Masica 1991: 103). Similarly, in Marathi, there is a tendency to spirantize postvocalic /b^h, d^h) and intervocalic /j^h/ (Lambert 1943). The phonetic differences between D^h and Δ almost negligible and phonologically unusable: there is no Indic language with both classes together, though Sindhi and Siraiki have another series of implosives) (cf. Masica 1991: 103; Ladefoged & Maddieson 1996: 57-60, 83). Generally, aspiration is very unusual with fricatives (and more often with sibilants than spirants). In Younger Avestan, there is a synchronic alternation between voiced fricatives and stops ($\Delta \sim D$); this process can

be both an innovation or a preservation of the older state. Especially for Indic is fortition of fricatives, even the synchronic process, affecting OIA sibilants, which regularly change to plosives before IE *s/š (see Appendix 3 for examples).

> **NB:** The development of clusters *TT, *TS is without any special features of interest for our current research.
>
> **NB:** De Angelis (2006) proposed that the paradigmatic pressure from Bartholomae´s Law forced all IE clusters of two voiced dentals *ddh to remain as such, with only two exceptions with inserted ᶻ, namely dehí, dhehí (< *dˇdh). However, our model does not require any inserted sibilant within the two-dental clusters, assuming the first plosive spirantization in both Bartholomaean and non-Bartholomaean clusters.

5. Conclusions

We therefore conclude with the following statements:

i. the progressive direction of the Bartholomaean Law is not in opposition to the regular regressive development of plosives: the voiced spirants were subject to their assimilation process that requires each adjacent plosive of a voiced spirant, either from left or from right, to be assimilated to a voiced spirant. The Bartholomaean outcomes are then not exceptions in the regular regressive development of plosives since plosives do not trigger them, although they were affected them;

ii. both Indic and Iranian outcomes with voiced plosives are later innovations resulting from a later parallel fortition, not the inherited Indo-Iranian phonemes. The Indic outcomes with voiced and voiced aspirated plosives are not results of a further dissimilation; the left spirant was never ʿaspiratedʾ, hence could not lose an aspiration. The Iranian outcomes with a single voiced plosive are also the later innovation; the attested existence of voiced spirants as allophones of voiced plosives in Avestan can be an indirect proof of the older existence of voiced spirants;

iii. suppose IE *Dʰ was a voiced spirant and not an aspirated voiced plosive (a phoneme which does not phonetically exist as such!). In that case, we do not have to assume the existence of any exotic *DZʰ clusters as Bartholomaean outcomes in the Common Indo-Iranian period both in IE dental and palatovelar series. The aspirated sibilants do exist, cf. Amerindian (mutually unrelated) Siouan language Ofo (/sʰ/ and /fʰ/), and in the (central?) Chumash languages (/sʰ/ and /ʃʰ/) usually voiceless (Sʰ). There is no need to shift an aspiration and create only positional aspirated sibilant if there is no aspiration at all;

iv. the process of fricative fortition, though relater rare, is well documented in Indic for IE sibilants (*s/š) and palatovelars (probably this is also the reason why IE *Tt is always OIA Tt, unlike other Indo-European languages).

Summarizing: the exceptional development of Bartholomae's Law is not unique at all if we consider the development of its parallel clusters. There were no shifts of aspiration or creation of specific and exotic voiced aspirated sibilants; the whole process is just a part of the development of clusters containing voiced spirants without regard to their relative *right–left* position.

Appendix i: the Indic outcomes

Table 4

Indic	*t-	*s-	*dʰ-
-*t/d	tt	ts	(d)dʰ
-*dʰ	**ddʰ**	*ts*	ddʰ
-*ḱ/ǵ	ṣṭ	kṣ	(d)dʰ
-*ǵʰ	**0ḍʰ**	*kṣ*	0dʰ
-*k/g	kt	kṣ	gdʰ
-*gʰ	**gdʰ**	*kṣ*	†gdʰ
-*p/b	pt	ps	†bdʰ
-*bʰ	**bdʰ**	*ps*	†bdʰ

NB: Bartholomaean outcomes are marked in **bold**. Outcomes due to a secondary levelling of originally Bartholomaean clusters are marked *cursive*. Directly unattested outcomes are marked by †. Examples of clusters resulting from *T/Dt*, *T/Ds* are willingly omitted.

Examples of Bartholomae's clusters (*Dʰt, *Dʰs):

pr. *inddhé*, ppp. *iddhá-* (√*idh-* 'kindle'; cf. pr. *indháte*, *indhaté*; < IE *√H_2eid^h-; cf. Gr. αἴθω 'I kindle'; cf. Whitney 1885: 8; MacDonell 1916: 371-372; IEW: 11-12; EWAi II: 267; LIV²: 259);

ppp. *baddhá-*, gd. *baddhvā́*, ft. *bhantsyáti* (√*bandh-* 'bind'; cf. pr. *badhnáte*; < IE *√b^hend^h-; cf. Goth. *band* 'bond'; cf. Whitney 1885: 105; MacDonell 1916: 400; IEW: 127; EWAi II: 208; LIV²: 75);

ppp. *buddhá-*, ao. *ábhutsi*, fut. *bhotsáti* B (√*budh-* 'wake'; cf. pr. *bódhati*; < IE *√b^heud^h-; cf. Gr. πεύθομαι 'I give notice', OCS *bljudǫ* 'beware'; cf. Whitney 1885: 106-107; MacDonell 1916: 400-401; IEW: 150-152; EWAi II: 233-235; LIV²: 82-83; NIL: 36-37);

ppp. *yuddhá-*, gd. *yuddhvī́*, pr. *yótsi*, des. *yúyutsati* (√*yudh-* 'fight'; cf. pr. *yúdhyate*; < IE *√H_ieud^h-; cf. OL *ioubē* 'I command'; cf. Whitney 1885: 133; MacDonell 1916: 410; IEW: 511-512; EWAi II: 418-419; LIV²: 225-226).

pr. *léḍhi* B (√*lih-* 'lick'; cf. caus. *leháyati*; < IE *√$leiǵ^h$-; cf. Gr. λείχω, OCS *ližǫ* 'I lick'; cf. Whitney 1885: 148; MacDonell 1916: 414; IEW: 668; EWAi II: 463; LIV²: 404);

ppp. *rūḍhá-*, gd. *rūḍhvā́*, ao. *rukṣās*, *árukṣat*, fut. *rokṣyáti* B, ds. *rúrukṣati* (√*ruh-* 'ascend'; cf. pr. *róhati*; < IE *√H_1leud^h-; cf. Gr. Hom. ἤλῦθον 'I came; cf. Whitney 1885: 143--144; MacDonell 1916: 414; IEW: 306-307, 684-685; EWAi II: 467-469; LIV²: 248--249; NIL: 245-246);

ppp. *ūḍhá-*, inf. *vóḍhum*, ao. *ávākṣur*, fut. *vakṣyáti* (√*vah-* 'carry'; cf. pr. *váhati*; < IE *√$ueǵ^h$-; cf. L *uehō*, OCS *vezǫ* 'I carry'; cf. Whitney 1885: 157; MacDonell 1916: 417; IEW: 1118-1120; EWAi II: 535-537; LIV²: 661-662);

ppp. *rabdhá-*, ds. *rípsate* B (√*rambh-* 'grasp'; cf. pr. *rábhate*; < IE *√$lemb^h$-; cf. Gr. λάφῦρα 'spoils of war'; cf. Whitney 1885:136; MacDonell 1916: 411; IEW: 652; EWAi II: 434-435; LIV²: 411-412); etymologically the same root as the following one;

ppp. *labdhá-*, gd. *labdhvā́*, ds. *álapsata*, fut. *lapsyáti*, des. *lípsate* AV, ps. *lipsyáte* B (√*labh-* 'take'; cf. pr. *lábhate*; < IE *√*lembʰ-*; cf. Gr. λάφῡρα 'spoils of war'; cf. Whitney 1885: 145-146; MacDonell 1916: 414; IEW: 652; EWAi II: 434-435; LIV²: 411-412);

ppp. *dabdhá-*, ds. *dípsati* (√*dabh-* 'harm'; cf. pr. *dábhati*; < IE *√*dʰebʰ-*; cf. Hitt. *tepnuzzi* 'downsize', Lith. *dóbiu* 'invalidate'; cf. Whitney 1885: 70; MacDonell 1916: 388; IEW: 240; EWAi II: 694-696; LIV²: 132-133; NIL: 85-86).

prc. *dhaktám* (√*dagh-* 'reach to'; cf. pr. *daghnuyā́t* ₆; < IE *√*dʰeugʷʰH₂-*; cf. Gr. Hom. φθάνω 'I come first'; cf. Whitney 1885: 69; MacDonell 1916: 388; IEW: 250; EWAi I: 691; LIV²: 134-135);

ppp. *dagdhá-*, ao. *ádhākṣīt*, fut. *dhakṣyáti* (√*dah-* 'burn'; cf. pr. *dáhati*; < IE *√*dʰegʷʰ-*; cf. Lith. *degù* 'I burn'; cf. Whitney 1885: 71; MacDonell 1916: 388-389; IEW: 240--241; EWAi I: 712-713; LIV²: 133-134);

pr. *dógdhi*, *dugdhé*, ao. *ádhukṣata*, *dhukṣata* (√*duh-* 'milk'; cf. pr. *duhánti*; < IE *√*dʰeugʰ-*; cf. Gr. τεύχω 'I make ready', OE *ge-dȳgan* 'overcome'; cf. Whitney 1885: 76; MacDonell 1916: 390; IEW: 271; EWAi I: 747-748; LIV²: 148-149).

Examples of parallel (non-Bartholomaean's) clusters (*T/DDʰ*, *DʰDʰ*):

pr. *addhí* (√*ad-* 'eat'; cf. pr. *ádmi*; < IE *√*H₁ed-*; cf. Hitt. *ēdmi*, L *edō* 'I eat'; cf. Whitney 1885: 3; MacDonell 1916: 370; IEW: 287-289; EWAi I: 61-62; LIV²: 230-231; NIL: 208-220);

pr. *viddhí* (√*vid-* 'find'; cf. pr. *vindáti*; < IE *√*u̯ei̯d-*; cf. L *uīdī* 'saw', Arm. *egit* 'found'; cf. Whitney 1885: 159-160; MacDonell 1916: 418; IEW: 1125-1127; EWAi II: 579--581; LIV²: 665-667; NIL: 717-722);

pr. *dehí*,[18] *daddhí* (√*dā-* 'give'; cf. pr. *dádāti*; < IE *√*deH₃-*; cf. Gr. δίδωμι, L *dō* 'I give', OLith. *duosti* 'give'; cf. Whitney 1885: 71-72; MacDonell 1916: 388-389; IEW: 223--225; EWAi I: 713-715; LIV²: 105-106; NIL: 60-69);

pf. *didiḍḍhí* (√*diś-* 'point'; cf. pr. *diśátu*; < IE *√*dei̯ḱ-*; cf. L *dīcō* 'I say'; cf. Whitney 1885: 73; MacDonell 1916: 389; IEW: 188-189; EWAi II: 744-746; LIV²: 108-109);

int. *nenigdhí* (√*nij-* 'wash'; cf. ao. *ánijam*; < IE *√*nei̯gʷ-*; cf. Gr. νίζω 'I wash', OIr. *-nenaig* 'wash'; cf. Whitney 1885: 90; MacDonell 1916: 395; IEW: 761; EWAi II: 54; LIV²: 450; NIL: 519-520; NIL: 660-661);

ao. *áyugdham* (√*yuj-* 'join'; cf. pr. *yuñjánti*; < IE *√*i̯eu̯g-*; cf. L *iungō* 'I harness', OCS *igo* 'yoke'; cf. Whitney 1885: 132-133; MacDonell 1916: 410; IEW: 508-510; EWAi II: 417-418; LIV²: 316; NIL: 397-404);

pr. *indhvám* (*-ddʰ-)[19] (√*idh-* 'kindle'; cf. pr. *indháte*; < IE *√*H₂ei̯dʰ-*; cf. Gr. αἴθω 'kindle'; cf. Whitney 1885: 8; MacDonell 1916: 371-372; IEW: 11-12; EWAi II: 267; LIV²: 259);

pr. *bodhí* (*-ddʰ-) (√*budh-* 'wake'; cf. pr. *bódhati*; < IE *√*bʰeu̯dʰ-*; cf. Gr. πεύθομαι 'I give notice', OCS *bljudǫ* 'I am beware'; cf. Whitney 1885: 106-107; MacDonell 1916: 400-401; IEW: 150-152; EWAi II: 233-235; LIV²: 82-83; NIL: 36-37);

---

[18] From *-d+dʰ*, output is simplified instead of the expected *daddʰi* or *dedʰi* (= *-0dʰ-*).

[19] However, clusters *nCC* of this type are often simplifed independently.

ao. *yodhí* (*-dd^h-) ($\sqrt{}$*yudh-* 'fight'; cf. pr. *yúdhyate*; < IE *$\sqrt{}$*H̥i̯eu̯d^h-*; cf. OL *ioubē* 'I command'; cf. Whitney 1885: 133; MacDonell 1916: 410; IEW: 511-512; EWAi II: 418--419; LIV²: 225-226);

pr. *dhehí*[20] ($\sqrt{}$*dhā-* 'put'; cf. pr. *dádhāmi*; < IE *$\sqrt{}$*d^heH₁-*; cf. Gr. τίθημι 'put', OLith. *dest*(*i*); cf. Whitney 1885: 82; MacDonell 1916: 392-393; IEW: 235-239; EWAi I: 783-786; LIV²: 136-138; NIL: 99-117);

ao. *voḷhám, voḍhvám* ($\sqrt{}$*vah-* 'carry'; cf. pr. *váhati*; < IE *$\sqrt{}$*u̯eǵ^h-*; cf. L *uehō*, OCS *vezǫ* 'I carry'; cf. Whitney 1885: 157; MacDonell 1916: 417; IEW: 1118-1120; EWAi II: 535-537; LIV²: 661-662).

Appendix ii: the Iranian outcomes

Table 5

Iranian	*t-	*s-	*d^h-
-*t/d	st	0s	zd
-*d^h	**zd**	**0z**	zd
-*ḱ/ǵ	št	0š	žd
-*ǵ^h	**žd**	**0ž**	†žd
-*k/g	xt	xš	gd
-*g^h	**gd**	**γš**	†gd
-*p/b	pt[21]	fš	bd
-*b^h	**bd**	**βž**	†bd

NB: Directly not attested outcomes are marked in *cursive*. The non-Bartholomaean outcomes parallel to the Bartholomaean are not listed. Directly not attested outcomes are marked by †. In the examples, outcomes of clusters resulting from **T/Dt*, **T/Ds* are willingly omitted again.

Examples of Bartholomae's clusters (**D^ht*, **D^hs*):

pr. OAv. *dazdā* but without Bartholomae's Law: YAv *dasta* ($\sqrt{}$*dā-* 'put'; cf. impf. YAv. *ādadaṯ*; < IE *$\sqrt{}$*d^heH₁-*; cf. Gr. τίθημι 'I put', OLith. *desti* 'put'; cf. IEW: 235-236; Kellens 1995: 29; LIV²: 136-137; Cheung 2007: 45-46);

---

[20] From -*d^h+d^h*, output is simplified instead of the expected *dadd^hi* or *ded^hi* (= -0*d^h-*).

[21] This outcome is unusual and unexpected, since we can expect †*ft*. It is a matter of debate, if attested *pt* is an archaism or re-archaization. The Avestan development of the cluster *Pt* is irregular, not only in comparison to the development of the analogous clusters but also in other Iranian languages, since Av. *hapta* 'seven' has the Pahlavi and New Persian counterpart *haft* (< OIranian *hafta). The question is whether Avestan *pt* is an archaism, or whether it represents an innovation (despirantization/occlusivization), since, in Avestan, we encounter the following forms of *pitar-* 'father': nom. sg. OAv. *ptā, tā*, YAv. *ptā, pita*, dat. sg. OAv. *fəδrōi, pisrē*, YAv. *pisre*. The form *tā* is easy to be explained as the result of development from *ftā (i.e., *pt > ft > ht > 0t*) but Hoffmann and Forssman (1996: 94) assume the direct simplification of the word-initial *pt-* > *t-*). Beekes (1988: 73) and Hoffmann & Forssman (1996: 94) otherwise reckon with the preservation of inherited *pt*, contrary to Reichelt (1909: 40), whereas for eastern dialects of Iranian Kümmel (2007: 65) assumes partial restitution of the spirants *f, *θ, *χ by the aspirates *p^h*, *t^h*, k^h, which partially, however, can be considered as original according to Morgenstierne (1942; also cf. Lipp 2009a: 158-160 with further references).

ppp. OAv. *vərəzda-* (√*vard-* 'grow', OAv. pr. *varədaitī*; < IE *√*H_ueRd^h*-; cf. OIA *várdhate* 'they grow'; cf. IEW: 1167; Kellens 1995: 51; LIV²: 228; Cheung 2007: 208);

ppp. OAv. *bazda-* but without Bartholomae's Law: YAv. *basta-* (√*band-* 'bind'; cf. ps. YAv. *baṇdaiiatati*; < IE *√*b^{h}end^h*-; cf. OIA *bandhati*, Goth. *bindan* 'bind'; cf. Kellens 1995: 37; LIV²: 75-76; Cheung 2007: 4-6);

ppp. YAv. *niuruzda-* but without Bartholomae's Law: YAv. *urusta-* (√*raud-* 'grow bigger'; cf. pr. *raoδahe*; < IE *√*$H_l$leu̯d^h*-; cf. Gr. Hom. ἤλὔθον 'I came'; cf. IEW: 306-307, 684-685; Kellens 1995: 59; LIV²: 248-249; NIL: 245-246; Cheung 2007: 193-194);

(?) pr. YAv. *uruuāza-* (√*uruuād-* 'be proud'; no secure IE cognate; cf. Kellens 1995: 60; Cheung 2007: 438);

pr. OAv. *gərəždā* (√*garz-* 'complain, pr. OAV. *gərəzōi*; < IE *√*g$^{(u)}$eRǵ^h*-; cf. OIA *gr̥hate* 'he complains'; cf. IEW: 350-351; Kellens 1995: 19-20; LIV²: 187; Cheung 2007: 111-112);

subs. OAv *vazdra-*, YAv. *važdri-* but without Bartholomae's law: ppp. YAv. *vašta-*, subs. *vaštar-* 'drag animal' (?)+ ao. YAv. *uzuuažat̰* (√*vaz-* 'drive', pr. YAv. *vazaiti*; < IE *√*u̯eǵ^h*-; < IE *√*u̯eǵ^h*-; cf. L *uehō*, OCS *vezǫ* 'I carry'; cf. IEW: 1118-1119; Kellens 1995: 52; LIV²: 661-662; Cheung 2007: 429-432);

pr. OAv. *aogᵊdā* but. pr. YAv. *aoxte* (√*aoj-* 'say'; cf. pr. OAv. *aojōi*; < IE *√*$H_l$eu̯g$^{(u)h}$*- (?); cf. Gr. εὔχομαι 'I pray', L *uoueō* 'I vow'; cf. IEW: 348; Kellens 1995: 9; Cheung 2007: 169-170);

subs. YAv. *dugəd(a)r-*, YAv. *duγδ(a)r-* 'daughter' (< IE *d^{h}ugH₂ter-;²² cf. OIA *duhitar-*, Gr. θὔγάτηρ, Goth. *daúhtar* 'daughter'; cf. IEW: 277; NIL: 126-130);

ds. OAv. *dīdraγžō.duiiē* (√*draṇj-* 'fix'; cf. caus. *drəṇjaiieiti*; < IE *√*dreg^h*-; cf. Gr. δρᾱσσομαι 'I grasp', OCS *drъžǫ* 'I hold'; cf. IEW: 254; Kellens 1995: 32; LIV²: 126; Cheung 2007: 76);

pr. OAv. *paⁱrii-aoγžā* (√*aoj-* 'say'; cf. pr. OAv. *aojōi*; < IE *√*H_lu̯eg$^{(u)h}$*- (?); cf. Gr. εὔχομαι 'I pray', L *uoueō* 'I vow'; cf. IEW: 348; Kellens 1995: 9, 14; LIV²: 253; Cheung 2007: 169-170);

subs. YAv. *dərəβδa* 'bundle of muscles' (< Iranian *√*darb-* 'join'; < IE *√*d^{h}erb^h*-; cf. Lith. *dirbù* 'I work'; cf. IEW: 211-212, 257; LIV²: 121; Cheung 2007: 60);

ds. OAv *diβžaidiiāi* (√*dab-* 'deceive'; cf. pr. OAv. *dəbənaotā*; < IE *√*d^{h}eb^h*-; cf. Lith. *dóbiu* 'subdue'; cf. IEW: 240; Kellens 1995: 27; LIV²: 132-133; NIL: 85-86; Cheung 2007: 42-43);

subs. YAv. *vaβžaka-* 'scorpion' (< IE *u̯ob^{h}sā*; cf. L *uespa*, OHG *wafsa*; 'wasp' cf. IEW: 1179).

Examples of parallel (non-Bartholomaean's) clusters (*T/DD^h, *D^{h}D^h):

pr. imp. YAv. *dazdi*, OAv. *mązdazdūm* (√*dā-* 'give'; cf. pr. OAv. *dadē*; < IE *√*deH₃*-; cf. Gr. δίδωμι, L *dō* 'I give', OLith. *duosti* 'give'; cf. IEW: 223-225; Kellens 1995: 29; LIV²: 105-106; NIL: 60-69; Cheung 2007: 43-45);

---

²² From IE *gH₂t*, probably merged with *g^{h}t* already before the Common Indo-Iranian period, since the first plosive was deaspirated according to Grassmann's Law, hence IE *gH₂t > *g^{h}t*.

inf. OAv. *vōizdiiāi*, ao. OAv. *frauuōizdūm* (√*vid-* 'know'; cf.. pf. OAv *vaēdā*; < IE *√*u̯eid̯-*; cf. L *uīdī* 'I saw', Arm. *egit* 'find'; cf. IEW: 1125-1127; Kellens 1995: 54; LIV^2: 665-667; NIL: 717-722; Cheung 2007: 408-409);

inf. YAv. *dazdiiāi* (√*dā-* 'put'; cf. impf. YAv. *ādadaṱ*; < IE *√*dʰeH₁-*; cf. Gr. τίθημι 'I put', OLith. *desti* 'put'; cf. IEW: 235-236; Kellens 1995: 29; LIV^2: 136-137; NIL: 99-11; Cheung 2007: 45-46);

inf. OAv. *āždiiāi* (√(*n*)*as/š-* 'reach'; cf. ao. OAv. *nąsaṱ*; < IE *√*H₂neḱ-*; cf. Gr. διηνεκής 'continuous'; cf. IEW: 316-317; Kellens 1995: 40-41; LIV^2: 282-283; Cheung 2007: 183-184);

inf. OAv. *mərąždiiāi* (√*marz-* 'rub'; cf. pr. OAv. *marəzaiti*; < IE *√*H₂merǵ-*; cf. Gr. ἀμέργω 'I pluck, pull'; cf. IEW: 722-723; Kellens 1995: 44; LIV^2: 280-281; Cheung 2007: 180-182);

pr. OAv. *mərəngəduiiē* (√*marc-* 'destroy'; cf. pr. *mərəṇcaite*; < IE *√*melkᵘ-*; cf. Gr. βλάπτω 'I damage'; cf. IEW: 737; Kellens 1995: 43; LIV^2: 434-435; Cheung 2007: 265-266);

pr. YAv. *auuaŋhabdaēta*, caus. YAv. *nixᵛabdaiieiti* (√*xᵛap-* 'sleep', pf. YAv. *hušxᵛafa* < IE *√*su̯ep-*; cf. L *sopiō*, OCS *sъpljǫ* 'I sleep'; cf. IEW: 1048-1149; Kellens 1995: 17-18; LIV^2: 612-613; NIL: 675-680; Cheung 2007: 145-146).

Appendix iii: the development of the **SDʰ* clusters in Indic

That Indic was subject to a regular de-spirantization is evident from the development of clusters **SDʰ* > OIA *0Dʰ*/*DDʰ* (cf. OIA imp. pr. *edhí* = √*as-* 'be' + *dhi*, OIA imp. pr. *ādhvam* = √*ās-* 'sit' + *dhvam*; OIA imp. pr. *viviḍḍhi* = √*viṣ-* 'be active' + *dhi*, OIA imp. pr. *rīḍhvam* = √*riṣ-* 'be hurt' + *dhvam*), since root-final sibilant was either lenited (de-buccalized and elided) or became a plosive in the position before a voiced plosive. Similarly, in clusters of two sibilants, the root-final usually became a plosive (cf. OIA ds. *jíghatsati* = √*ghas-* 'eat' + *sa-*, OIA *ávatsyat* в. = √*vas-* 'shine' + *sya-*; OIA inj. ao. *dvikṣát* = √*dviṣ-* 'hate' + *sá-*; OIA pr. *vívekṣi* = √*viṣ-* 'be active' + *si*).[23]

Avestan preserved the original left sibilant, which was lost before a following sibilant, this state of arts is an archaic feature, replaced by a new creation in Indic.

Trajectories we assume for the development of these clusters are:

ia. S + Dʰ > ZΔ > i̯Dʰ   (Indic)
ib. S + Dʰ > ZΔ > ΔΔ > DDʰ   (Indic)
ii. S + Dʰ > ZΔ > ZD   (Iranian)

i. S + S > Θs > TS   (Indic)
ii. S + S > SS > 0S   (Iranian)

---

[23] OIA *ási* from √*as-* is probably the old IE simplification of clusters, cf. L. *es*. OIA *śāssi* = √*śās-* + *si* is a unique example of the (restituted?) cluster.

Examples:
Indic:

pr. *edhí* (√*as-* 'be'; cf. pr. *ásat*; < IE *√*H₁es-*; cf. Gr. ἐστί, L *est* 'is'; cf. Whitney 1885: 5; MacDonell 1916: 370-371; IEW: 340-341; EWAi I: 144; LIV²: 241-242; NIL: 235- -238);

pr. *ā́dhvam* (√*ās-* 'sit'; cf. pr. *ā́sāthe*; < IE *√*H₁eH₁s-*; cf. Hith. ēsa, Gr. ἧσται 'sits'; cf. Whitney 1885: 6-7; MacDonell 1916: 371; IEW: 342-343; EWAi II: 181; LIV²: 232);

pr. *śādhí*, pf. *śaśādhí* (√*śās-* 'order'; cf. pr. *śā́smi*; < IE *√*ḱeHs-*; cf. Alb. *thom* 'say'; cf. Whitney 1885: 172; MacDonell 1916: 423; IEW: 533; EWAi II: 632-633; LIV²: 318- -319);

pf. *viviḍḍhí* (√*viṣ-* 'be active'; cf. pr. *víveṣaḥ*; < IE *√*u̯eis-* 3; cf. Whitney 1885: 161; MacDonell 1916: 419; EWAi II: 585-586; LIV²: 672);

ao. *rīḍhvam* TA (√*riṣ-* 'be hurt'; cf. pr. *ríṣyati*; < IE *√*(H₁)reis-*; cf. Whitney 1885: 140; IEW: 859; EWAi II: 462; LIV²: 505);

ds. *jíghatsati* AV (√*ghas-* 'eat'; cf. pf. *jaghā́sa*; < IE *√*g⁽ʷ⁾hes-*; cf. OAv. *aγžōnuuamna-* 'not consumable'; cf. Whitney 1885: 42; MacDonell 1916: 381; IEW: 452; EWAi I: 514; LIV²: 198-199);

co. *ávatsyat* B (√*vas-* 'shine'; cf. ao. *ávasran*; < IE *√*u̯es-*; cf. Lith. *aũšti* 'break dawn'; cf. Whitney 1885: 155-156; MacDonell 1916: 417; IEW: 86-87; EWAi II: 530-532; LIV²: 292-293; NIL: 357-367);

ao. *dvikṣát*, *dvikṣata* AV (√*dviṣ-* 'hate'; cf. pr. *dvéṣat*; < IE *√*du̯eis-*; cf. OAv. *dᵃⁱbišəṇtī*; cf. Whitney 1885: 81; MacDonell 1916: 392; IEW: 228; EWAi I: 770-771; LIV²: 131);

ao. *apikṣan* ŚB (√*piṣ-* 'crush'; cf. pf. *pipéṣa*; < IE *√*peis-*; cf. Lith. *pisù* 'I copulate', OCS *pьxotъ* 'I pushed'; cf. Whitney 1885: 97-98; MacDonell 1916: 398; IEW: 796; EWAi II: 169; LIV²: 466-467);

pr. *vívekṣi*, fut. *vekṣyáti* (√*viṣ-* 'be active'; cf. pr. *víveṣaḥ*; < IE *√*u̯eis-* 3; cf. Whitney 1885: 161; MacDonell 1916: 419; EWAi II: 585-586; LIV²: 672).

Iranian:

pr. OAv. *zdī* (√*ah-* 'be'; cf. pr. OAv. *ahmī*; < IE *√*H₁es-*; cf. Gr. ἐστί, L *est* 'is'; cf. IEW: 340-341; Kellens 1995: 10-11; LIV²: 241; Cheung 2007: 151-152);

ao. OAv. *θrāzdūm* (√*θrā-* 'protect'; cf. YAv. pr. *θrāiieṇte*; < IE *√*treH-*; cf. OIA *trā́yate* 'he protects, saves'; cf. IEW: 1075; Kellens 1995: 27; LIV²: 646; Cheung 2007: 394);

inf. OAv. *sazdiiiāi*[24] (√*səⁿh-* 'declare', pr. YAv. *saŋhaite*; < IE *√*ḱens-*; cf. L *cēnseō* 'I judge'; cf. IEW: 566; Kellens 1995: 62; LIV²: 326; Cheung 2007: 334-335);

ao. OAv. *θβarōždūm* (√*θβars-* 'cut'; cf. pr. YAv. *θβərəsaiti*; < IE *√*tu̯ers-* (?); cf. IEW: 1102; Kellens 1995: 26; LIV²: 656; Cheung 2007: 399-400);

pr. (ao.?) OAv. *cīždī* (√*ciš-* 'assign'; cf. pr. OAv. *cīšmahī* < IE *√*kʷeis-*; cf. OIr. *ad-cí* 'see'; cf. IEW: 637; Kellens 1995: 22-23; LIV²: 381-382; Cheung 2007: 30);

inf. OAv. *būždiiāi* (√*būš-* 'endeavour'; < IE *√*bʰeu̯H₂-s-*; cf. Lith. *bùs* 'it will be'; cf. IEW: 146-147; Kellens 1995: 39-40; LIV²: 98-101; Cheung 2007: 25-26);

---

[24] This from can be from √*saṇd-* 'appear", cf. Cheung (2007: 334), but it seems to be doubtful. Kellens lists it as derived from √*səⁿh-* 'declare" without any doubts (Kellens 1995: 62).

part. inch. YAv. (*vī*)*usaitī*- f. (√*vah*- 'shine', < IE *√*u̯es*-; cf. Lith. *aũšti* 'break dawn'; cf. IEW: 86-87; Kellens 1995: 53; LIV²: 292-293; Cheung 2007: 202);

ppp. (?) YAv. *ustriiamna*- (= *us*-√*stər*- 'throw down'; < IE *√*ster*-; cf. L *prosternō* 'cause to fall'; cf. IEW: 1029-1030; Kellens 1995: 64; LIV²: 597-598; Cheung 2007: 363-364);

pr. inch. YAv. *tusən* (√*tuš*- 'be empty'; cf. pr. caus. YAv. -*taošaiieiti*; < IE *√*teu̯s*-; cf. OCS *tъštь* 'empty, vain', Lith. *tùščias* 'empty, poor'; cf. IEW: 1085; Kellens 1995: 26; LIV²: 642; Cheung 2007: 388-389).

Abbreviations

ao.	= aorist		inch.	= inchoative
co.	= conditional		inf.	= infinitive
ds.	= desiderative		int.	= intensive
fut.	= future		pf.	= perfect
gd.	= gerund		pr.	= present
imp.	= imperative		ppp.	= preterit passive participle
impf.	= imperfect		subs.	= substantive

Languages

Alb.	= Albanian		OAv.	= Old Avestan
Arm.	= Armenian		OCS	= Old Church Slavonic
Av.	= Avestan		OE	= Old English
Goth.	= Gothic		OIA	= Old Indo-Aryan
Gr.	= Greek		OIr.	= Old Irish
Hitt.	= Hittite		OL	= Old Latin
Hom.	= Homeric		OLith.	= Old Lithuanian
IE	= Indo-European		OP	= Old Persian
L	= Latin		YAv.	= Young Avestan
Lith.	= Lithuanian			

Symbols

D	= voiced unaspirated plosive		T	= voiceless unaspirated plosive
Dʰ	= voiced aspirated plosive		Tʰ	= voiceless aspirated plosive
Δ	= voiced spirant		Θ	= voiceless spirant
S	= voiceless sibilant		Z	= voiced sibilant
Sʰ	= voiceless aspirated sibilant		Zʰ	= voiced aspirated sibilant

References

Anderson, Stephen R. 1970. On Grassmann's Law in Sanskrit. *Linguistic Inquiry* 1(4). 387-396.

Bartholomae, Christian. 1882. Die arische Vertretung von Med. Asp. + *t* und Med. Asp. + *s*. In Bartholomae, Christian. *Arische Forschungen I.*, 3-24. Halle: Niemeyer.

Bartholomae, Christian. 1883. *Handbuch der altiranischen Dialekte*. Leipzig: Breitkopf & Härtel.

Bartholomae, Christian. 1885. Θυγάτερ. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der indogermanischen Sprachen* (*KZ*) 27 (Neue folge VII), 206-207.

Bartholomae, Christian. 1895-1901. I. Vorgeschichte der iranischen Sprachen. II. Awestasprache und Altpersisch. In Geiger, Wilhelm & Kuhn, Ernst (eds), *Grundriß der iranischen Philologie*, Erster Band. I. Abteilung, 1-248. Strassburg: Karl J. Trübner.

Beekes, Robert S. P. 1988. *A grammar of Gatha-Avestan*. Leiden: Brill.

Blust. Robert. 2009. *The Austronesian languages*. Canberra: Pacific Linguistics. Research School of Pacific and Asian Studies–The Australian National University.

Brixhe C. 1997. Un "nouveau" champ de dialectologie grecque: Le macédonien. In Cassio, Albio C. (ed.), *Katà diálekton: Atti del III Colloquio Internazionale di Dialettologia Greca. Napoli – Fiaiano d'Ischia, 25-28 settembre 1996.* AIΩN XIX, 41-71. Napoli: Istituto Universitario Orientale.

von Brücke, Ernst. 1856. *Grundzüge der Physiologie und Systematik der Sprachlaute* für *Linguisten und Taubstummenlehrer*. Wien: Carl Gerold's Sohn.

Brust, Manfred. 2018. *Historische Laut- und Formenlehre des Altpersischen – mit einem etymologischen Glossar* (Innsbrucker Beiträge zur Sprachwissenschaft 161). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, Bereich Sprachwissenschaft.

Cheung, Johny. 2007. *Etymological dictionary of the Iranian verb*. Leiden: Brill.

Collinge, Neville E. 1985. *The laws of Indo-European*. Amsterdam–Philadelphia: John Benjamins.

De Angelis, Alessandro. 2006. A. ind. *dhehí, dehí* e la legge di Bartolomae come processo di "terapia linguistica". *Archivio glottologico italiano*. 91(1). 3-30.

Ejerhed, Eva I. 1981. The analysis of aspiration in Sanskrit phonology. *Nordic Journal of Linguistics* 4(2). 139-159.

EWAi = Mayrhofer, Manfred. 1992. *Etymologisches Wörterbuch des Altindoarischen I.* + Mayrhofer, Manfred. 1996. *Etymologisches Wörterbuch des Altindoarischen II.* + Mayrhofer, Manfred. 2001. *Etymologisches Wörterbuch des Altindoarischen III*. Heidelberg: Carl Winter.

Görtzen, Jens. 1998. *Die Entwicklung der indogermanischen Verbindungen von dentalen Okklusiven mit besonderer Berücksichtigung des Germanischen* (Innsbrucker Beiträge zur Sprachwissenschaft94)*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Hammerich, Louis L. 1967. Ketzereien eines alten Indogermanisten. In *To honor Roman Jakobson: Essays on the occasion of his 70. birthday, 11. October 1966*, vol. 2, 839-849. The Hague: Mouton.

Hill, Eugen. 2003. *Untersuchungen zum inneren Sandhi des Indogermanischen: Der Zusammenstoß von Dentalplosiven im Indoiranischen, Germanischen, Italischen und Keltischen*. Bremen: Hempen Verlag.

Hoffmann, Karl & Forssman, Bernard. 1996. *Avestische Laut- und Flexionslehre*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Hopper, Paul. 1973. Glottalized and murmured occlusives in Indo-European. *Glossa* 7. 141-166.

IEW = Pokorny, Julius 1959*. Indogermanisches etymologisches Wörterbuch. I. Band, A-KN*. Bern: Francke Verlag; 1966. *Indogermanisches etymologisches Wörterbuch. II. Band, KO-WR*. Bern: Francke Verlag. 1969. *Indogermanisches etymologisches Wörterbuch. III. Band.* Bern: Francke Verlag.

Jakobson, Roman. 1958. Typological studies and their contribution to historical comparative linguistics. In Sievertsen, Eva (ed.), *Proceedings of the Eighth International Congress of Linguists,* 17-35. Oslo: Oslo University Press.

Kellens, Jean. 1995. *Liste du verbe avestique*. Wiesbaden: Dr. Ludwig Reichert Verlag.

Kobayashi. Masato. 2004. *Historical phonology of Old Indo-Aryan consonants*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa. Tokyo University of Foreign Studies.

Kümmel, Martin J. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen* für die *vergleichende Rekonstruktion*. Wiesbaden: Reichert.

Kümmel, Martin J. 2012. *Typology and reconstruction*. In Nielsen Whitehead, Benedicte & Olander, Thomas & Olsen, Birgit A. & Rasmussen, Jens E. (eds.), *The sound of Indo-European*, 291-329. Copenhagen: Museum Tusculanum Press.

Kuryłowicz, Jerzy. 1935. *Études indoeuropéennes I*. Kraków: Skład główny w księgarni Gebethnera i Wolffa.

Ladefoged, Peter & Maddieson, Ian. 1996. *The sounds of the world's languages*. Malden MA: Blackwell Publishers.

Lambert, Hester M. 1943. *Marathi language course*. Calcutta–Oxford: Humphrey Milford.

Lehmann, Winfred P. 1952. *Proto-Indo-European phonology*. Austin: The University of Texas Press–Linguistic Society of America.

Lipp, Reiner. 2009a. *Die indogermanischen und einzelsprachlichen Palatale im Indo-iranischen.* Band I: *Neurekonstruktion, Nuristan-Sprachen, Genese der indoarischen Retroflexe, Indoarisch von Mitanni.* Heidelberg: Universitätsverlag Winter.

Lipp, Reiner. 2009b. *Die indogermanischen und einzelsprachlichen Palatale im Indo-iranischen.* Band II: *Thorn-Problem, indoiranische Laryngalvokalisation.* Heidelberg: Universitätsverlag Winter.

LIV² = Rix, Helmut & Kümmel, Martin J. & Zehnder, Thomas & Lipp, Reiner & Schirmer, Brigitte. 2001. *Lexikon der indogermanischen Verben.* Wiesbaden: Dr. Ludwig Reichert Verlag.

Lombardi, Linda. 1991. *Laryngeal features and laryngeal neutralization.* University of Massachusetts. (Doctoral dissertation.)

Lubotsky, Alexander. 2018. The phonology of Proto-Indo-Iranian. In Klein, Jared & Joseph, Brian & Fritz, Matthias (eds.), *Handbook of comparative and historical Indo-European linguistics* (Handbooks of Linguistics and Communication Science 41.3), vol. 3, 1876-1888. Berlin–Boston: De Gruyter Mouton.

MacDonell, Arthur A. 1910. *Vedic grammar.* Strassburg: Karl J. Trübner.

MacDonell, Arthur A. 1916. *A Vedic grammar for students.* Oxford: Clarendon Press.

Masica, Colin. P. 1991. *The Indo-Aryan languages.* Cambridge: Cambridge University Press.

Mayrhofer, Manfred. 1986. *Indogermanische Grammatik I.2. Lautlehre.* Heidelberg: Winter.

Mayrhofer, Manfred. 2004. *Die Hauptprobleme der indogermanischen Lautlehre seit Bechtel.* Wien: Verlag der österreichischen Akademie der Wissenschaften.

Mey, Jacob L. 1972. Was Bartholomae really a Grassmann? *Norwegian Journal of Linguistics* 1. 81-89.

Miller, D. Garry. 1977. Bartholomae's Law and an IE root structure constraint. In Hopper, Paul J. (ed.), *Studies in descriptive and historical linguistics: Festschrift for W. P. Lehmann*, 365-392. Amsterdam: John Benjamins.

Morgenstierne, Georg. 1942. Orthography and sound-system of the Avesta. *Norsk Tidsskrift for Sprogvidenskap* 12. 30-82.

NIL = Wodtko, Dagmar S. & Irslinger, Britta & Schneider, Carolin. 2008. *Nomina im indogermanischen Lexikon.* Heidelberg: Winter Universitätsverlag.

Prokosch, Eduard. 1918-1919. Media Aspirata (I-IV). *Modern Philology* 15(10). 621-628; 16(2). 99-115; 16(6). 325-336; 16(10). 543-552.

Prokosch, Eduard. 1939. *A comparative Germanic grammar.* Philadelphia: Linguistic Society of America – University of Pennsylvania.

Reichelt, Hans. 1909. *Awestisches Elementarbuch.* Heidelberg: Carl Winter's Universitätsbuchhandlung.

Sag, Ivan A. 1974. The Grassmann's Law ordering pseudoparadox. *Linguistic Inquiry* 5(4). 591-607.

Sag, Ivan A. 1976. Pseudosolutions to the pseudoparadox: Sanskrit diaspirates revisited. *Linguistic Inquiry* 7(4). 609-622.

Schindler, Jochem. 1976. Diachronic and synchronic remarks on Bartholomae's Law and Grassmann's Law. *Linguistic Inquiry* 7(4). 622-637.

Šefčík, Ondřej. 2012. On the integration of Old Indo-Aryan voiceless aspirates into the system. In Sukač, Roman & Šefčík, Ondřej (eds), *Sound of Indo-European 2*, 271-280. München: Lincom Europa.

Szemerényi, Oswald. 1990. *Einführung in die vergleichende Sprachwissenschaft.* Darmstadt: Wissenschaftliche Buchgesellschaft.

Walde, Alois. 1897. Die Verbindungen zweier Dentale und tönendes *z* im Indo-germanischen. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der indogermanischen Sprachen* (*KZ*) 34 (Neue Folge 17). 461-536.

Whitney, William D. 1885. *The roots, verb-forms and primary derivatives of the Sanskrit language: A supplement to his Sanskrit grammar.* Leipzig: Breitkopf and Härtel.

First preliminary report on the project for a new Cushitic comparative dictionary

Václav Blažek[1] &[*] Gábor Takács[2]

[1]Department of Linguistics & Baltic Studies, Masaryk University, Brno
blazek@phil.muni.cz | ORCID: 0000-0002-6797-7188
[2]Department of Classical Philology, University of Łódź (UŁ)
Lexicographical Library of Afro-Asiatic Root Research at Balatonederics (LLAARR)
Member of the Associazione Internazionale di Studi sul Mediterraneo e l'Oriente, Roma (ISMEO)
gabtak@datatrans.hu | ORCID: 0000-0002-2466-6451

Abstract: The paper reports about the start of the joint research project by the two authors for the 4th attempt at composing the comparative dictionary of all Cushitic languages.

Keywords: comparative and historical linguistics, lexicography, linguistic reconstruction, etymology, Afro--Asiatic, Cushitic languages.

> Dedicated to the blessed memories of
> Aharon Dolgopolsky (1930-2012) and
> Marvin Lionel Bender (1934-2008),
> Authors of a few giant steps of
> Cushitic and Omotic lexical comparison

Introduction

The Cushitic languages, in the current classification,[1] represent the fourth branch of the immense Afro-Asiatic macrofamily beside (1) Semitic, (2) Egyptian, (3) Berber, (4) Omotic and (5) Chadic. According to several lexicostatistical-glottochronological

---

[*] This paper is to hallmark the 30th anniversary of the collaboration between both authors, whose correspondence on Afro-Asiatic etymologies has started in March 1994.

[1] Established by J.H. Greenberg (1955, 1963) and modified by M.L. Bender (1975) adding Omotic.

calculations, Cushitic has been multiply corroborated as one of the earliest branch to have split off from the Afro-Asiatic parental community.[2]

Since none of the Cushitic languages were written in ancient times and Cushitic lexical items were only very scarcely recorded in ancient Egyptian sources,[3] they are only known as spoken today or in the best case from the records of the past one and a half century made mostly by European lininguists, which is why the historical linguistic reconstruction of the Cushitic lexicon has approximately similar importance and chances as, e.g., in the Fenno-Ugric (Uralic) language family (first attested sporadically in medieval Hungarian).

The *communis opinio* has divided Cushitic in four equipotential subbranches: (1) northern (Bed'awye of the nomadic Beja tribes located along the Red Sea coastline from Erithrea up to Egypt), (2) central (Agaw in Ethiopia), (3) East Cushitic (the biggest one, spread across Ethiopia, Somalia, Kenya), (4) Southern Cushitic (where several daughter languages are either extinct or endangered on the border of Kenya and Tanzania, the southernmost extremity of the entire macrofamily).

Previous Cushitic lexicons

So far three common Cushitic lexicons of a similar scope, albeit fundamentally diverging in their approach and nature of their methodology, have been published.

The pioneering milestone was set by A.B. Dolgopol'skij (Moscow), the co-founder mastermind of the Muscovite school of historical linguistics, whose ground-breaking *Sravnitel'no-istoričeskaja fonetika kušitskih jazykov* ("Comparative-historical phonology of the Cushitic languages", usually abbreviated SIFKJa, where Omotic was included also as not yet treated as a separate branch at that day) has appeared in 1973 after a hole decade of comparative research on the Cushitic lexicon (starting in the early 1960s), where he has first demonstrated a system of regular consonantal correlations through hundreds of reconstructed Cushomotic lexical items, whose principal weakness is that, following the convention since L. Reinisch, the Viennese giant of Ethiopian field research and the supreme authority of Cushomotic linguistics for many decades, numerous distinct roots were forced together in one entry, while Dolgopol'skij immediately targeted to set up Proto-Cushomotic forms without first establishing the corresponding forms of the intermediate phases at the younger diachronic levels, e.g., Proto-Agaw vs. Proto-East Cushitic or Proto-Somaloid vs. Proto-Highland East Cushitic. Later on, esp. after his

---

[2] Already I.M. D'jakonov (1975) has pointed out by lexicostatistical means Cushitic (handled at that day yet as one branch with Omotic) to have branched off PAA as the first back in the 8th mill. BC.

[3] Cf., e.g., the exceptionally ancient occurence (probably the oldest one at all in the history Cushitic, of the Somali term for "frankincense" in the Egyptian texts as ʕntj.w (earliest attestation in Dyn. IV, cf. ÄWb I 279b, a special innovative formation of Somali, which is nowhere else attested in Cushitic, from ECu. *ʕan- "milk"), which was discovered by L. Reinisch (1902: 61 and 93) and then thoroughly demonstrated by G. Banti and R. Contini (1997: 175, 184), cf. also Takács (2014, coll. 8b-9b). Another old Cushitic trace ("only" some half of a millennium younger than the above one) has been detected among the dog names in the Middle Egyptian inscription of king Antef II (of Theban Dyn. XI, ca. 2100 BCE), cf. Kossmann 2011 (with further literature).

aliyah (1976), Dolgopolsky himself abandoned his 1973 proto-forms and exploited the new results in reconstructing the Cushitic subbranches by H.-J. Sasse (1979) and Ch. Ehret (1980) for his milestone studies establishing some regular consonantal correspondences between East Cushitic and Egypto-Semitic (Dolgopolsky 1983 and 1988) as well as Southern Cushitic and Egypto-Semitic (Dolgopolsky 1987).

The circumstances for this latter approach have substantially bettered by the time when Ch. Ehret (UCLA) launched the second enterprise at a "Proto-Cushitic Reconstruction" (1987), which has greatly benefited from the epoch-maker progress in reconstructing the parental phonology and lexicon of Agaw by D. Appleyard (1984), of East Cushitic by P. Black (1974) and especially by H.-J. Sasse (1979) and of Southern Cushitic by Ch. Ehret (1980), although the latter work has rightly been blamed for little semantical rigour in establishing cognate pairs, which makes precisely the South Cushitic column of his comparative lexicon (where Omotic was excluded) less reliable and the weakest piece of chain in the whole work, which otherwise challenges to be used with care even today.

The third attempt was done by M.L. Bender somewhen between 2003-7,[4] but the outcome had had to be long awaited (posthumously ed. by G. Hudson in 2020), where Bender followed the lexostatistic strategy of identifying cognacy without an emphasis on setting up too many complicated common proto-forms which in the indices suggest an utterly minimalistic conception of the parental Cushitic consonantal inventory. He, as usually in other comparative lexicons of his also, was working with selected semantical items of (mostly) the basic lexicon for achieving cognate pairs and their consonantal correlations and so a great deal in the vocabularies of the compared languages do not show up, while the entries are arranged in the alphabetic order of the selected English meanings.

Needless to say that all three attempts at Proto-Cushitic greatly differ in reconstructional methodology and, as a consequence, their results also. Each tool has its pros and cons and their progressive features complement one another. What makes the first attempt unchallenged in this context is Dolgopol'skij's unique collection of comparative data (which only sometimes has to be re-arranged) and the true (not distorted or re-formatted) presentation of all the forms quoting directly the authentic sources with author names, while his original vision, first formulated in his SIFKJa, about a refined system of Cushomotic sibilant affricates (that must be modified within Cushitic vs. Omotic on several points) has underlied the new model of Afro-Asiatic phonology[5] as demonstrated in the

---

[4] Which can be dated from the completion of his Omotic comparative lexicon (2003) until his premature passing away (in Jan. 2008), after which G. Hudson managed to overtake and edit Bender's practically complete manuscript.

[5] A.Ju. Militarev and O.V. Stolbova (1990: 45), successors of Dolgopol'skij in the Diakonovian dictionary team in a way, have expressed their view on this original vision of "the phonological principles of Afrasian reconstruction" in their talk on their "First approach to comparative-historical phonology of Afrasian (consonantism)" delivered for the Viennese Afro-Asiatic Congress (1987) as follows: "The funny thing about them is that A. Dolgopolsky's brilliant hypothesis on Proto-Cushitic consonantism, he does not seem to insist upon any more, were among stimulating initial impulses in our work. What was not always well sustained in Cushitic has paradoxically proved more promising for Common Afrasian."

revolutionary Afro-Asiatic comparative vocabulary (SISAJa I-III) by the team of I.M. D'jakonov a decade later. The highly fruitful exploitation of the Cushitic sublevel (Agaw, East and South Cushitic) reconstructions as well as the daring conception of the Cushitic parental consonantism (equally figured with a few sibilant affricates, only otherwise as in SIFKJa) make Ehret 1987 even today an unparalleled rich (over 640 entries with no semantical restrictions) and attractive inventory of Proto-Cushitic roots which approached the closest towards the maximally possible score of his day in this respect (it is only a pity that in 3 columns out of the 4, i.e., in Agaw, East and South Cushitic, mostly asterisked were only quoted, sometimes difficult to check back without the underlying real forms in the daughter languages). The comparative analysis of select-ed (mostly) basic semantemes in by Bender (2020) Cushitic has yielded reliable bases for certain lexicostatistical and glottochronological calculations instead of a comprehen-sive inventory of all the Cushitic roots one might possibly isolate.

The new comparative Cushitic project

Regarding the enumerated achievements as well as deficiencies of the three pioneering works, without which our present agenda would be unthinkable, one can presumably easily imagine the requirements we are to state as most fundamental *sine qua non* about the new Cushitic comparative dictionary:

(1) First and foremost, it should represent an as complete and encyclopaedic thesaurus of all the lexical items that occur at least in two Cushitic subbranches and are not orig-inating from some recent borrowing (e.g., from Ethio-Semitic), whereby they may be reliably used for recunstructing an older common Cushitic heritage.

(2) In an appendix, in addition, all the lexemes should also be presented that only occur (for the time being) in one subbranch but do not appear to be a loanword and so that these may also be exploited in later interbranch comparison.

(3) All the lexical data should be quoted in their authentic forms, referring to the sources where these may be checked back, with the necessary comments (if needed) as for orthographical, transcriptional peculiarities or phonetic realization.

(4) The compared forms should adhere to regular consonantal correspondences whose establishment may reliably only be elaborated and achieved by a careful comparison of the essence drawn from the results by the three predecessors with our own research re-sults. In each individual etymological entry, any change deviant from the expectable correspondence should be remarked and possibly commented on in special (foot)note. The lexical material should be arranged according to the smaller units headed preferably by the adequate sublevel proto-form. If the sporadic attestation of the compared forms does not make the cognacy evident, additional elucidation on the supposed intermediate phonological shifts is required. In case of an alternative etymology, special comment should be adduced in a footnote for the introduction and a brief discussion of the prob-lem, which should result in an annotated apparatus attached to each entry or disputed form. Each etymological entry has to begin with a Proto-Cushitic form to which all previous suggestions by other authors should be adduced in a footnote.

As an outcome, an A-Z thesaurus, representing the whole of the immense Cushitic lexical heritage in thoroughly annotated and footnoted entries, is expected by both co-authors, namely Václav Blažek (VB) and Gábor Takács (GT), who have shared the domains of their responsability as follows:

Bed'awye or Beja

In the parts of Beja Etymological Dictionary VB has done his best for an exhaustive documentation from the beginning of the 19th cent. till the present (Blažek 2003a, 2003b, 2005a, 2006, 2007, 2013b, 2020, 2021, 2023). VB supposes that, for a comprehensive tool like our Cushitic Comparative Dictionary, it is not that practical to quote all records for the same item and so VB would chose the best transcribed examples, e.g. 2-3, with full documentation of their sources covering the same item.

Agaw

Although D. Appleyard has basically established Proto-Agaw over some two decades of publishing his research results (1984-2006), this was in fact achieved by comparing the lexical evidence for certain semantical items from selected daughter languages as recorded by selected field researchers, complemented by VB with additional Agaw daughter languages and sources on the basis of a 100 wordlist for lexicostatistical purposes (Blažek 2006-2007a; 2014: 299-315), which GT intends to complete by all missing semantical items and available records as well as emend the previous comparisons on certain points in the manner of his pilot study (Takács 2012).

East Cushitic

Concerning the East Cushitic part of the Cushitic comparative dictionary, VB has planned a maximum transparent pattern organized according to individual subgroups: 0. East Cushitic protoform, I. Afar-Saho, II. Somaloid, III. Galaboid, IV. Oromoid, V. Dullay, VI. Highland East Cushitic, VII. Yaaku. VB would like to quote all adequate examples from all described languages, always with sources. In case of multiple sources for the same lexeme, the more precisely transcribed example should be preferred. In the case of semantic dispersion in various sources, all examples from one and the same languages should be quoted (frequently it is the case of "big" languages as Oromo, Somali, Afar-Saho, Sidamo with richer documentation). It was Paul Black in his dissertation (1974) who initiated the application of the standard comparative method to the East Cushitic languages. Thanks to his field research, he had at his disposal fresh data from Konso and D'irayta, two archaic Oromoid languages. His results were developed by Hans-Jürgen Sasse in a series of articles (1973, 1974, 1975, 1976, 1979) and crowned by his *Etymological dictionary of Burji* (1982), where he significantly expanded the

number of reliable Proto-East Cushitic reconstruction. Let us mention that Sasse newly described Dasanech (= Galab) and Burji languages. He also had an experience in diachronic Indo-European linguistics and was able to apply its principles to the comparative East Cushitic linguistics. In Sasse's *Etymological dictionary of Burji*, he included numerous comparisons identified by Dick Hayward, whose contribution to East Cushitic comparative linguistics was much more extensive as he published both descriptive and comparative studies devoted to the Dullay group (1978, 1989), Bayso (1978-9), D'irayta (1981), Afar (1985, co-authored by E. Parker) and Arbore (1984, 1988). Another scholar who shifted the level of reconstruction in the field the East Cushitic languages, namely of the group 'Sam' called by the author himself, was Bernd Heine (1976, 1978, 1981, 1982). Extraordinarily important are his comparative descriptions of Elmolo (1980) and especially Yaaku (1975). Challenged by Ch. Ehret & M. N. Ali's (1984) Proto-Somali based on the comparative survey of the Somali dialect for certain lexical items, the very useful reconstruction of Proto-Somali Marcello Lamberti (1986) was based on the same languages as Heine's Proto-Sam was (Somali dialects, Jiddu, Boni, Rendille). His daring alternative theory on labiovelar correlations in Cushitic-Omotic (Lamberti 1988) is, however, rather hard for us to follow. Grover Hudson (1989) reconstructed a considerable part of the Highland East Cushitic protolexicon, including a rich lexical documentation according to English semantical items, collecting all possible forms for these items from the daughter languages including those not underlying any of the reconstructed items. Chris Ehret (1991) proposed his own conception of reconstructing Proto-East Cushitic phonetics, preceded by his attempt to reconstruct the Common Cushitic protolanguage (1987). He tried to expand the number of ejective consonants in both his East and Common Cushitic reconstructions. These ideas of Ehret have been further developed by Linda Arvanites in her dissertation (1991), devoted to the Proto-East Cushitic reconstruction. Taking into account all these attempts at reconstructing East Cushitic, we prefer the approach of Hans-Jürgen Sasse as the most convincing and reliable one. In addition, VB has published several studies mapping the selected partial (basic) lexicons of some individual East Cushitic groups (Somaloid, Dullay, Oromoid etc.) in a comparative context (2001, 2006-2007b, 2010, 2011; with Jan Záhořík 2008).

Southern Cushitic

No doubt, Ch. Ehret's (1980) pioneering phonological and lexical reconstruction of this southernmost, lexically pretty intact, subbranch of the entire Afro-Asiatic phylum requires fundamental revision for long known[6] methodological reasons. This unique vocabulary, whose 100-item comparative wordlist has been re-arranged by VB on the basis of Ehret 1980 along with other sources for lexicostatistical purposes (Blažek 2005b; 2019: 45-88), has been since 1998 under permanent reconstruction by GT, who has released since then a whole series of papers and volume chapters on the exceptionally archaic traits of Southern Cushitic consonantism with a great number of newly reconstructed

---

[6] Cf., e.g., the assessment by R. Hetzron & E.P. Tálos (1982).

lexical items (Takács 1999-2011). In addition, over the past many years, an entirely re-written and re-arranged Southern Cushitic comparative root dictionary has been on schedule by GT based on a comprehensive survey and re-evaluation of all the available lexical materials, to become an etymological dictionary, with due respect to the achievements so far in the comparison and reconstruction of the best studied West Rift languages (Whiteley 1958, Elderkin & Maghway 1992, Kießling 2002, Kießling & Mous 2004), which are being re-examined in constant contrast with East Rift, Ma'a and Dahalo (the other well-descrived lexicon of the subbranch). Although GT is aware of the recent dispute over the classification of the latter two languages, before drawing any hasty conclusions, he maintains the exploitation of their lexical treasure definitely displaying characteristic features of Southern Cushitic or Rift especially regarding the fine distinction of the labial triad (*b-, *p-, *f-) and their laterals (extinct from East Cushitic).

References

Appleyard, D. 1984. The internal classification of the Agaw languages: A comparative and historical phonology. In Bynon, J. (ed.), *Current progress in Afro-Asiatic Linguistics*, 33-67. Amsterdam–Philadelphia: John Benjamins.

Appleyard, D. 1991. The vowel systems of Agaw: Reconstruction and historical inferences. In *Proceedings of the Fifth International Hamito-Semitic Congress*, vol. II, 13-28. Wien: Afro-Pub.

Appleyard, D. 1996. The position of Agaw within Cushitic. In Zemánek, P. (ed.), *Studies in Near Eastern languages and literatures: Memorial volume for Karel Petráček*, 1-14. Praha: Academy of Sciences of the Czech Republic, Oriental Institute.

Appleyard, D. 1996. 'Kaïliña' – a 'new' Agaw dialect and its implications for Agaw dialectology. In Hayward, R.J. & Lewis, I.M. (eds.), *Voice and power: The culture of language in North-East Africa*, 1-19. London: SOAS.

Appleyard, D. 1996. Preparing a comparative Agaw dictionary. In Griefenow-Mewis, C. & Voigt, R. (eds.), *Cushitic and Omotic languages: Proceedings of the Third International Symposium. Berlin, March 1994*, 185-200. Köln: Rüdiger Köppe Verlag.

Appleyard, D. 2006. *A comparative dictionary of the Agaw languages*. Köln: Rüdiger Köppe Verlag.

Arvanites, Linda. 1991. *The glottalic phonemes of Proto Eastern Cushitic*. University of Calfornia (UCLA): Los Angeles. (Doctoral dissertation).

ÄWb I = Hannig, R. 2003. *Ägyptisches Wörterbuch I. Altes Reich und Erste Zwischenzeit* (Hannig-Lexica 4). Mainz am Rhein: Verlag Philipp von Zabern.

Banti, Giorgio & Contini, Riccardo.1997. Names of aromata in Semitic an Cushitic languages. In Avanzini, A. (ed.), *Profumi d'Arabia: Atti del convegno* (Saggi di storia antica 11), 169-192. Roma: «L'Erma» di Bretschneider.

Bender, M.L. 1975. *Omotic: A new Afroasiatic language family*. Carbondale, Illinois: Southern Illinois University.

Bender, Marvin Lionel. 2020. *Cushitic lexicon and phonology*. Edited by Grover Hudson. (Schriften zur Afrikanistik / Research in African Studies 28). Berlin etc.: Peter Lang.

Black, P.D. 1974. *Lowland East Cushitic: Subgrouping and Reconstruction*. New Haven: Yale University. (Doctoral dissertation).

Blažek Václav. 1997. Cushitic lexicostatistics: The second attempt. In *Afroasiatica Italiana. Studi Africanistici* (Seria Etiopica 6), 171-188. Napoli: Istituto Universitario Orientale.

Blažek, Václav. 2001. Rendille lexicon – state of the art. *Afrikanistische Arbeitspapiere* 65. 255-264.

Blažek, Václav. 2003a. Fauna in Beja lexicon. In Kogan, Leonid (ed.), *Orientalia: Trudy Instituta vostočnyh kul'tur. III: Studia Semitica (Fs. A. Militarev)*, 230-294. Moskva: Rossijskij Gosudarstvennyj Gumanitarnyj Univerzitet.

Blažek, Václav. 2003b. Beja kinship and social terminology. In Hasitzka, Monika R.M. & Diethart, Johannes & Dembski, Günther *Das alte Ägypten und seine Nachbarn: Festschrift zum 65. Geburtstag von Helmut Satzinger*, 307-340. Krems: Österreichisches Literaturforum.

Blažek, Václav. 2005a. Natural phenomena, time and geographical terminology in Beja lexicon (Fragment of a Comparative and Etymological Dictionary of Beja), I. Kogan, Leonid E. & Koslova, Natalia & Loesov, Sergey & Tishchenko, Serguei (eds.), *Memoriae Igor M. Diakonoff* (Babel und Bibel 2), 365-407. Winona Lake, IN: Eisenbrauns.

Blažek, Václav. 2005b. Current progress in South Cushitic Comparative Historical Linguistics. *Folia Orientalia* 42. 177-224.

Blažek, Václav. 2006. Natural phenomena, time and geographical terminology in Beja lexicon (Fragment of a Comparative and Etymological Dictionary of Beja), II. In Kogan, Leonid E. & Koslova, Natalia & Loesov, Sergey & Tishchenko, Serguei (eds.), *Annual of Ancient Near Eastern, Old Testament, and Semitic Studies* (Babel und Bibel 3), 383-428. Winona Lake, IN: Eisenbrauns.

Blažek, Václav. 2006-2007a. A new study of an Agaw language (on *The Kemantney language*, by Z. Leyew, Köln: Köppe 2003). *Folia Orientalia* 42/43. 353-367.

Blažek, Václav. 2006-2007b. Current progress in Cushitic linguistics: Ts'amakko (on *A grammar of Ts'amakko*, by G. Savà, Köln: Köppe 2005). *Folia Orientalia* 42/43. 369-388.

Blažek Václav. 2007. Beja historical phonology: Consonantism. In Amha, Azeb & Mous, Maarten & Savà, Graziano (eds.), *Omotic and Cushitic language studies: Papers from the Fourth Cushitic Omotic Conference (Leiden, April 2003)*, 125-156. Köln: Köppe.

Blažek, Václav. 2010. Glottochronological classification of Oromo dialects. *Lingua Posnaniensis* 52(2). 27-42.

Blažek, Václav. 2011. Dasanech language – state-of-the-art. *Folia Orientalia* 48. 267-279.

Blažek, Václav. 2013a. Cushitic year. In Mengozzi, Alessandro & Tosco, Mauro (eds.), *Sounds and words through the ages: Afroasiatic studies from Turin*, 65-79. Alessandria: Edizioni dell'Orso,.

Blažek, Václav. 2013b. Beja lexicostatistics. *Lingua Posnaniensis* 55. 21-35.

Blažek, Václav. 2013c. Afroasiatic migrations. In Ness, Immanuel & Bellwood, Peter (eds.), *The encyclopedia of global human migration*, vol. I, 125-132. Oxford: Wiley-Blackwell.

Blažek, Václav. 2014. On classification of Agaw languages. *Folia Orientalia* 51. 295-318.

Blažek, Václav. 2019. South Cushitic classification in lexicostatistic perspective. *Folia Orientalia* 56. 41-94.

Blažek, Václav. 2020. An excerpt from the "Comparative and etymological dictionary of Beja: Anatomical lexicon". *Folia Orientalia* 57. 25-125.

Blažek, Václav. 2021. Flora in Beja Lexicon. *Folia Orientalia* 58. 39-63.

Blažek, Václav. 2023. An excerpt from the "Comparative and etymological dictionary of Beja: Natural phenomena, time and geographical terminology". *Folia Orientalia* 60. 329-428.

Blažek, Václav & Záhořík, Jan. 2008. Bayso: An endangered language of the Ethiopian highlands. In Machalík, T. & Mildnerová, K. & Záhořík, J. (eds.)**,** *Viva Africa: Proceedings of the 3rd International Conference on African Studies*, 53-70. Plzeň: Adéla.

D'jakonov, I.M. 1975. Lingvističeskie dannye k istorii drevnejših nositelej afrazijskih jazykov. *Africana. Afrikanskij Étnografičeskij Sbornik* 10. 117-130.

Dolgopol'skij, A.B. 1973. *Sravnitel'no-istoričeskaja fonetika kušitskih jazykov*. Moskva: Nauka.

Dolgopolsky, A. 1983. Semitic and East Cushitic: Sound correspondences and cognate sets. In Segert, S. & Bodrogligeti, A.J.E. (eds.), *Ethiopian studies dedicated to Wolf Leslau*, 123-142. Wiesbaden: Otto Harrassowitz.

Dolgopolsky, A. 1987. South Cushitic lateral consonants as compared to Semitic and East Cushitic. In Jungraith-mayr, H. & Müller, W.W. (eds.), *Proceedings of the Fourth International Hamito-Semitic Congress*, 195-214. Amsterdam: John Benjamins.

Dolgopolsky, A. 1988. Semitic and East Cushitic: Word-initial laryngeals. In Taddese, B. (ed.), *Proceedings of the Eighth International Conference of Ethiopian Studies, University of Addis Ababa, 1984*, vol. 1, 629-637. Addis Ababa: Institute of Ethiopian Studies, Addis Ababa.

Ehret, Ch. 1980. *The historical reconstruction of Southern Cushitic phonology and vocabulary*. Berlin: Dietrich Reimer Verlag.

Ehret, Ch. 1987. Proto-Cushitic reconstruction. *Sprache und Geschichte in Afrika* 8. 7-180.

Ehret, Ch. 1991. The consonant inventory of Proto-Eastern Cushitic. *Studies in African Linguistics* 22(3). 211-275.

Ehret, Ch. & Ali, M. N. 1984. Soomaali classification. In Labahn, T. (ed.), *Proceedings of the Second International Congress of Somali Studies*, vol. 1, 201-269. Hamburg: Buske Verlag.

Elderkin, E.D. & Maghway, J.B. 1992. Some West Rift roots. *African Languages and Cultures* 5/1. 43-64.

Greenberg, J.H. 1955. *Studies in African linguistic classification*. Branford, Connecticut: Compass Publishing Company.

Greenberg, J.H. 1963. The languages of Africa. *International Journal of American Linguistics* 29.

Hayward, Dick (Richard J.). 1978. The Qawko dialects and Yaaku. *Abbay* 9. 59-70.

Hayward, Dick (Richard J.) 1978. Bayso revisited: Some preliminary linguistic observations. I. *Bulletin of the School of African and Oriental Research* 46(3). 540-570.

Hayward, Dick (Richard J.). 1979. The place of Bayso within Eastern Cushitic. In Hess, R.L. (ed.), *Proceedings of the Fifth International Conference on Ethiopian Studies. Session B. April 13-16, 1978, Chicago, USA*, 73-90. Chicago: University of Chicago Press.

Hayward, Dick (Richard J.). 1979. Bayso revisited: Some preliminary linguistic observations. II. *Bulletin of the School of African and Oriental Research* 47(1). 101-132.

Hayward, Dick (Richard J.). 1981. Nominal suffixes in Dirayta (Gidole). *Bulletin of the School of Oriental and African Studies* 44. 126-144.

Hayward, Dick. (= Richard. J.). 1984. *The Arbore language: A first investigation including a vocabulary*. Hamburg: Helmut Buske Verlag.

Hayward, Dick (Richard J.): Comparative Notes on the Language of the S'amakko.= Journal of Afro-Asiatic Languages 1 (1989), 1-53.

Heine, B. 1975. Notes on the Yaaku language (Kenya). *Afrika und Übersee* 58(2). 119-138.

Heine, Bernd. 1976. Notes on the Rendille Language. *Afrika und Übersee* 59. 176-223.

Heine, Bernd. 1977. Bemerkungen zur Boni-Sprache (Kenia). *Afrika und Übersee* 60. 242-295.

Heine, B. 1978. The Sam languages: A history of Rendille, Boni and Somali. *Afroasiatic Linguistics* 6(2). 23-115.

Heine, Bernd. 1980. *The non-Bantu languages of Kenya*. Berlin: Reimer.

Heine, Bernd. 1980. Elmolo. In *The non-Bantu languages of Kenya* (Language and Dialect Atlas of Kenya 2), 173-218. Berlin: Reimer.

Heine, Bernd. 1981. Some cultural evidence on the early Sam-speaking people of Eastern Africa. *Sprache und Geschichte in Afrika* 3. 169-200.

Heine, B. 1982. *Boni dialects*. Berlin: Dietrich Reimer Verlag.

Hetzron, Robert & Tálos, E. Péter. 1982. Christopher Ehret's "The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary". A Review Article. *Sprache und Geschichte in Afrika* 4. 239-250.

Hudson, G. 1989. *Highland East Cushitic dictionary*. Hamburg: Buske.

Kießling, Roland. 2002. *Die Rekonstruktion der südkuschitischen Sprachen (West-Rift), von den systemlinguistischen Manifestationen zum gesellschaftlichen Rahmen des Sprachwandels* (Kuschitische Sprachstudien 19). Köln: Rüdiger Köppe Verlag.

Kießling, R. & Mous, M. 2004. *The lexical reconstruction of West-Rift Southern Cushitic* (Kuschitische Sprach-studien, Band 21). Köln: Rüdiger Köppe Verlag.

Kossmann, Maarten. 2011. The names of King Antef's dogs. In Mettouchi, A. (ed.), *«Parcours berbères»: Mélanges offerts à Paulette Galand-Pernet et Lionel Galand pour leur 90ᵉ anniversaire*, 79-84. Köln: Rüdiger Köppe Verlag.

Lamberti, M. 1986. *Die Somali-Dialekte: Eine vergleichende Untersuchung.* Hamburg: Helmut Buske Verlag.

Lamberti, M. 1988. The correspondence "labial-velar-glottal" in Cushitic. In Bechhaus-Gerst, M. & Serzisko, F. (eds.), *Cushitic-Omotic: Papers from the International Symposium on Cushitic and Omotic Languages, Cologne, January 6-9, 1986*, 303-308. Hamburg: Helmut Buske Verlag.

Militarev, A. & Stolbova, O.V. 1990. First approach to comparative-historical phonology of Afrasian (Consonantism). In Mukarovsky, H.G. (ed.), *Proceedings of the Fifth International Hamito-Semitic Congress*, Bd. I, 45-72. Wien: Afro-Pub.

Parker, E.M. & Hayward, R.J. 1985. *An Afar-English-French dictionary (with grammatical notes in English)*. London: School of Oriental and African Studies, University of London.

Reinisch, Leo. 1902. *Die Somali-Sprache*. II. *Wörterbuch*. Wien: Alfred Hölder Verlag.

Sasse, H.-J. 1973. Spuren von Pharyngalen im Galab. *Afrika und Übersee* 56. 266-275.

Sasse, H.-J. 1974. Notes on the structure of Galab. *Bulletin of the School of Oriental and African Studies* 37. 407-438.

Sasse, H.-J. 1975. Galla /š/, /s/ und /f/. *Afrika und Übersee* 58. 244-263.

Sasse, H.-J. 1976. Weiteres zu den ostkuschitischen Sibilanten. *Afrika und Übersee* 59. 125-142.

Sasse, H.-J. 1979. The consonant phonemes of Proto-East-Cushitic (PEC). *Afroasiatic Linguistics* 7(1). 1-67.

Sasse, H.-J. 1982. *An etymological dictionary of Burji*. Hamburg: Helmut Buske Verlag.

SIFKJa = Dolgopol'skij, A. B. 1973. *Sravnitel'no-istoričeskaja fonetika kušitskih jazykov*. Moskva: Nauka.

SISAJa I-III = D'jakonov, I.M. & Belova, A.G. & Četveruhin, A.S. & Militarev, A.Ju. & Porhomovskij, V.Ja. & Stolbova, O.V. 1981-1986. *Sravnitel'no-istoričeskij slovar' afrazijskih jazykov*. Vypusk 1-3. Pis'mennye pamjatniki i problemy istorii kul'tury narodov Vostoka. XV-XIX godičnaja naučnaja sessija Leningradsko-go Otdelenija Instituta Vostokovedenija Akademii Nauk SSSR. Moskva: Nauka.

Takács, G. 1999. Sibilant and velar consonants of South Cushitic and other Afro-Asiatic branches. In Lamberti, M. & Tonelli, L. (eds.), *Afroasiatica Tergestina. Papers from the 9th Italian Meeting of Afro-Asiatic (Hamito-Semitic) Linguistics, Trieste, April 23-24, 1998. Contributi presentati al 9o Incontro di Linguistica Afroasiatica (Camito-Semitica), Trieste, 23-24 Aprile 1998*, 393-426. Padova: Unipress.

Takács, G. 2000. South Cushitic consonant system in Afro-Asiatic context. *Afrikanistische Arbeitspapiere* 61. 69-117.

Takács, G. 2001. Towards Proto-Afro-Asiatic phonology: Ancient remnants in South Cushitic, Angas-Sura, and North Bauchi. *Rocznik Orientalistyczny* 54(2). 55-125.

Takács, G. 2002. Ma'a lexicon and Afro-Asiatic III. *Rocznik Orientalistyczny* 55(1). 109-133.

Takács, G. 2003. South Cushitic sibilant affricates in a comparative-historical perspective. In Bender, M.L. (chief ed.) & Appleyard, D. & Takács, G. (eds.), *Selected comparative-historical Afrasian linguistic studies in memory of Igor M. Diakonoff* (Lincom Studies in Afroasiatic Linguistics 14), 143-162. München & Newcastle: Lincom Europa.

Takács, G. 2005. Problems of Afro-Asiatic historical phonology: Ancient remnants of sibilant affricates in South Cushitic and Chadic. In Fronzaroli, P. & Marrassini, P. (eds.), *Proceedings of the 10th Meeting of Hamito-Semitic (Afroasiatic) Linguistics (Florence, 18-20 April 2001)* (Quaderni di Semitistica 25), 65-83. Firenze: Dipartimento di Linguistica, Università di Firenze.

Takács, G. 2005. Recent problems of Semitic-Egyptian and Semito-Cushitic and -Chadic consonant corre-spon-dences. In Olmo Lete, G. del (ed.), *Proceedings of the Barcelona Symposium on Comparative Semitic (19-20 November 2004)*, *Aula Orientalis* (Barcelona) 23(1-2). 207-231.

Takács, G. 2005. New progress in the reconstruction of Southern Cushitic: Review of Kießling, Roland: Die Rekonstruktion der südkuschitischen Sprachen (West-Rift), Von den systemlinguistischen Manifestationen zum gesellschaftlichen Rahmen des Sprachwandels (Kuschitische Sprachstudien Band 19, Köln, 2002., Rüdiger Köppe Verlag) and Kießling, Roland & Mous, Maarten: The Lexical Reconstruction of West-Rift Southern Cushitic (Kuschitische Sprachstudien Band 21, Köln, 2004., Rüdiger Köppe Verlag). *Lingua Posnaniensis* 47. 213-225.

Takács, G. 2009. Ma'a Lexicon and Afro-Asiatic IV: Ma'a ŝ-. *Lingua Posnaniensis* 51. 125-131.

Takács, G. 2009. South Cushitic sibilant affricates in a comparative-historical Afro-Asiatic perspective II. In Monti, A. & Minetti, F. & Gallucci, F. (eds.), *Eastern perspectives: From Qumran to Bollywood*, 135-142. Alessandria: Edizioni dell'Orso.

Takács, G. 2010. The Afro-Asiatic background of South Cushitic *ḫ, *ḥ, *ʕ, *h, and *ʔ. *Journal of Language Relationship* 4. 91-122

Takács, Gábor. 2011. *Studies in Afro-Asiatic comparative phonology: Consonants*. Berlin: Dietrich Reimer Verlag.

Takács, Gábor. 2012. Agaw lexicon and its Cushitic and Afro-Asiatic background. *Rocznik Orientalistyczny* 65(2). 85-118.

Takács, Gábor. 2014. Questions of the African lexicon in Older Egyptian. *Orientalistische Literaturzeitung* 109(4-5). 1-13.

Whiteley, W. H. 1958. *A short description of item categories in Iraqw (with material on Gorowa, Alagwa and Burunge)*. Kampala: East African Institute of Social Research.

Preliminary report on the project for an Etymological dictionary of the Angas-Sura group of Chadic languages

Gábor Takács

Department of Classical Philology, University of Łódź (UŁ)

Lexicographical Library of Afro-Asiatic Root Research at Balatonederics (LLAARR)

Member of the Associazione Internazionale di Studi sul Mediterraneo e l'Oriente, Roma (ISMEO)

gabtak@datatrans.hu | ORCID: 0000-0002-2466-6451

Abstract The brief report announces the current work on the forthcoming first two fascicles of an etymological dictionary elaborating the Afro-Asiatic background of a Chadic language group, one of the first ones in this genre in the history of Chadic linguistics.[1]

Keywords: African linguistics, Afro-Asiatic, Chadic, etymology.

Introduction

The languages of the Angas-Sura group, spoken between the South-Eastern Plateau and the Benue river, Plateau State of Nigeria. belong to the West Chadic subbranch of

---

[1] The primacy belonging, in a way at least, to N. Skinner's (1996) *Hausa comparative dictionary* (Cologne, 1996, Köppe) which, however, lists *comparanda* in and outside Afro-Asiatic in the manner of J.H. Greenberg's „mass comparison" *ad hoc*, without elaborating the consonantal correspondences. Others out of the 27 (!) Chadic groups have mostly not even been been researched even for an internal reconstruction, let alone for the elaboration of their etymological dictionaries wit all the Afro-Asiatic cognates etc. So far, the comparative lexicons and (sometimes) historical phonologies have only appeared for the following groups: Angas-Sura (Takács 2004) with pilot studies on their Afro-Asiatic background (see fn. 6 below), Bole-Tangale (Schuh 1984), Northern Bauchi (Skinner 1977 and Takács 2002) with some external etymologies by the present author (Takács 2002, 2007), Southern Bauchi (Shimizu 1978), Mafa-Mada (Rossing 1978), all in all just 5 (!) Chadic groups, while efforts have been made to globally reconstruct West Chadic (Stolbova 1986-7) and Central Chadic (Gravina 2014) also, though both attempts suffer from fundamental difficulties of coherence regarding their historical phonologies surely to be reconsidered in the light of additional data.

the Chadic branch, which, in its turn, represents part of the Afro-Asiatic (Semito-Hamitic) language family (or phylum), divided since the third quarter of the 20th century into six equipotential cognate branches thanks to the fundamental results of J. Lukas & J.H. Greenberg on Chadic[2] and M. L. Bender & H. C. Fleming on Omotic,[3] resp., namely: (1) Semitic, (2) Egyptian, (3) Berber, (4) Cushitic, (5) Omotic, (6) Chadic.

Some minor segments of the phonological and lexical reconstruction of the Angas-Sura group had only been partly elaborated in pioneering studies by J. H. Greenberg, C. Hoffmann and O. V. Stolbova[4] in the second half of the 20th century before the first comprehensive comparative lexicon of Angas-Sura has been published by G. Takács (2004)[5] that also only offers a purely internal reconstruction without going beyond the frontiers of Angas-Sura.

The new project

Over the past two decades, on the basis of that comparative dictionary, I have discussed thousands of external cognates of the Angas-Sura lexical stock both inside its gigantic Chadic kindred and in the remote branches of the Afro-Asiatic macrofamily in my series of papers "Angas-Sura etymologies", whose majority has appeared on the pages of this journal.[6] Using these materials with numerous *addenda et corrigenda* re-

---

[2] Who were the first scholars to claim the Chadic languages to belong together in a separate branch of Afro-Asiatic.

[3] The two American Ethiopianists were both the pioneers of Omotic comparative linguistics establishing their internal classification as well as the isomorphs and the isoglosses indicating their special status in contrast to the Cushitic languages which they were both the first ones to recognize.

[4] Thus, J.H. Greenberg (1958) surveyed the Angas-Sura roots beginning with labials pointing out the original labial triad *b - *p - *f inherited from Afro-Asiatic. O.V. Stolbova devoted two studies to the subject, using basically the Angas (Foulkes 1915, Ormsby 1913-4) and Sura (Jungraithmayr 1963) lexicons for the comparison adducing some additional data from Chip, Montol, Gerka (collected and published by Jungraithmayr 1965). In 1972, she proposed a historical-comparative survey of the Proto-Angas-Sura consonant system in the light of some illustrative lexical material (2-3 exx. for each correspondence). In her 1977 paper, O. V. Stolbova presented 256 lexical roots and Proto-Angas-Sura reconstructions accompanied by a brief sketch of vowel correspondences. C. Hoffmann (1975 MS) offered a phonological (both consonantal and vowel) reconstruction of the Proto-Angas-Goemai level (on the basis of Goemai, Mernyang, Sura, and Angas) through 248 lexical roots. The West Chadic historical phonology by Stolbova (1987: 240-244) also contains a separate list of some 64 Proto-Angas roots.

[5] My research on the Afro-Asiatic background of the Angas-Sura lexicon began back in September 1998 during my first research at the Haifa University under the guidance of the late Prof. A. B. Dolgopolsky (1930-2012), may his memory be blessed, one of the greatest Afro-Asiatic or Semito-Hamitic comparativists of all times.

[6] First part (roots with initial *b-) in *Lingua Posnaniensis* 46 (2004), 131-144. Second part (*ḅ-) in *Rocznik Orientalistyczny* 57/1 (2004), 55-68. Third issue (*p-) in *Lingua Posnaniensis* 48 (2006), 121--138. Fourth part (*f-) in *Folia Orientalia* 47/2 (2011), 273-289. Fifth part (*m- in monoconsonantal roots) in *Cahiers Caribéens d'Egyptologie* 13-14 (2010), 137-142. Sixth part (rest of *m-), originally scheduled for *Rocznik Orientalistyczny* 74/1 (2021), has so far not been completed and submitted. Seventh part (*d-) in *Lingua Posnaniensis* 62/3 (2020), 95-120. Eighth part (*ḍ-) in *Folia Orientalia* 57 (2020), 321-354. Ninth part (*t-) in *Lingua Posnaniensis* 63/1 (2021), 53-72. Tenth part (*z- + Ø or

sulting from most recent new sources on the Angas-Sura daughter languages as well as from my ongoing research on other Afro-Asiatic groups in the frames of my current project for the micro-reconstructions in Southern Afro-Asiatic,[7] I am currently working on the reformed etymological entries for the Angas-Sura lexical items with initial labials and dentals to achieve a new synthesis going into the first two separate volumes of the planned „Etymological dictionary of the Angas-Sura group of Chadic languages".[8] As my research for collecting comparative evidence with initial sibilant and velar consonants in my quarter-of-a-century old Afro-Asiatic root catalogue progresses these years, systematically dealing with further segments of the Angas-Sura lexical stock will also be facilitated for hopefully updating or elaborating all the relevent etymological entries for the next two fascicles.

References

Fitzpatrick, J.F.J. 1910-1911. Some notes on the Kwolla District and its tribes. *Journal of the Royal African Society* 10. 16-52, 213-22.

Foulkes, H.D. 1915. *Angass manual: Grammar, vocabulary*. London: Kegan Paul, Trench, Trübner and Co.

Gravina, R. 2014. *Central Chadic reconstructions*. (https://www.academia.edu/19594305/ Central_Chadic_Reconstructions) (Accessed 2014.)

Greenberg, J.H. 1958. The labial consonants of Proto-Afro-Asiatic. *Word* 14. 295-302.

Hoffmann, C. 1975. Towards a comparative phonology of the languages of the Angas-Goemai group. MS. University of Ibadan, faculty seminar on 19 March 1975. 32 p.

Jungraithmayr, H. 1963. Die Sprache des Sura (Maghavul) in Nordnigerien. *Afrika und Übersee* 47. 8-89, 204-220.

Jungraithmayr, H. 1965. Materialen zur Kenntnis des Chip, Montol, Gerka und Burrum (Südplateau, Nordnige-rien). *Afrika und Übersee* 48. 161-183.

Ormsby, G. 1912-1913 & 1913-1914. Notes on the Angass language. *Journal of the Royal African Society* 12 (1912-1913). 421-424 & 13 (1913-1914). 54-61, 204-210, 313-315.

labials or dentals or velars) in *Lingua Posnaniensis* 64/1 (2022), 73-96. Eleventh part (*z- + nasals) in *Lingua Posnaniensis* 64/2 (2022), 49-76. Twelfth part (*z- + liquids) in *Lingua Posnaniensis* 63/2 (2021), 56-75. Thirteenth parts with AS *z- + liquids (continued) in *Lingua Posnaniensis* 65/2 (2023), 83-102.

Rossing, M. O. 1978. *Mafa-Mada: A comparative study of Chadic languages in North Cameroun.* Wisconsin: University of Wisconsin (Madison). (Doctoral dissertation.)

Schuh, Russell G. 1984. West Chadic vowel correspondences. In Bynon, James (ed.), *Current progress in Afro-Asiatic Linguistics: Papers of the Third International Hamito-Semitic Congress* (London, 1978) (Current Issues in Linguistic Theory 28), 167-223. Amsterdam: Benjamins.

Shimizu, Kiyoshi. 1978. *The Southern Bauchi group of Chadic languages*. Marburg: *Africana Marburgensia*, Special Issue 2, 1-50.

Skinner, Neil. 1977. North Bauchi Chadic languages: Common roots. *Afroasiatic Linguistics* 4(1). 1-49.

Stolbova, O.V. 1972. Soglasnye verhnezapadnočadskih jazykov: Genetičeskie sootvetstvija. In Makaev, É.A. (ed.), *Istoriko-tipologičeskie i sinhronno-tipologičeskie issledovanija (na materiale jazykov raznyh sistem)*, 179--187. Moskva: Institut Jazykoznanija Akademii Nauk SSSR.

Stolbova, O.V. 1977. Opyt rekonstrukcii verhnezapadnočadskih kornej. In *Jazyki zarubežnogo Vostoka: Sbornik statej*, 152-160. Moskva: Nauka.

Stolbova, O. V. 1986. Rekonstrukcija konsonantnoj sistemy zapadnočadskih jazykov. In *Pis'mennye pamjatniki i problemy istorii kul'tury narodov Vostoka: XIX godičnaja naučnaja sessija Leningradskogo Otdelenija Instituta Vostokovedenija Akademii Nauk SSSR*, 80-115. Moskva: Nauka.

Stolbova, O.V. Sravnitel'no-istoričeskaja fonetika i slovar' zapadnočadskih jazykov.= Porhomovskij, V.Ja. (ed.): Afrikanskoe istoričeskoe jazykoznanie. Problemy rekonstrukcii. Moskva, 1987, Nauka. 30-268.

Takács, G. 2002. Outline of a North Bauchi historical phonology. *Afrikanistische Arbeitspapiere* 70. 167-190.

Takács, G. 2004. *Comparative dictionary of the Angas-Sura languages*. Berlin: Dietrich Reimer Verlag.

Takács, G. 2007. The Afro-Asiatic background of the North Bauchi consonant system II. In Tourneux, H. (ed.), *Topics in Chadic linguistics III: Historical studies. Papers from the 3rd Biennial International Colloquium onthe Chadic Languages, Villejuif, November 24-25, 2005*, 103-131. Köln: Rüdiger Köppe Verlag.

Kupchik, John 2023. *Azuma Old Japanese: A comparative grammar and reconstruction* (Trends in Linguistics. Documentation 40). Berlin–Boston: De Gruyter Mouton. Pp. XXXIV + 522

Alfred F. Majewicz

International Institute of Ethnolinguistic and Oriental Studies, Stęszew
majewicz@amu.edu.pl | ORCID: 0000-0002-8984-3148

There are three key reasons why the very cover and title page of the book under review[1] here caught this writer's eye: (1) the term *Azuma*, (2) the glottonym *Old Japanese*, and (3) the name of the author of the book. Actually, this writer has been familiar with all the three "lexical items", yet it was spotting them together on the cover of a volume just released that triggered not only the present text but also an irresistible impulsion to have the book at hand on the shelf of his own library along with standing there other works of primary importance focusing on, or related to, Old Japanese. Astonishingly (or rather *not* astonishingly[2]), the date of publication of almost all of these works starts with <20> and almost all of them are either authored (~ coauthored) by Alexander ("Sasha") Vovin or in some other way (acknowledgements of academic supervision or support, citations, references, polemics, etc.) associated with him.

In the second edition of his Western Old Japanese (WOJ) grammar (2020:20), he wrote that "[...] there are *very few general descriptions* of the" WOJ "language grammar published even in Japanese. Surprisingly enough, *none of these descriptions is complete*"[3]; he briefly characterized only one such grammar and three[4] "much shorter works" (*ib.*) to

---

[1] Addressed, in the first place, to "general linguists" who are interested in results presented in such works, but not engaged or specializing in, or unrelated to, Japanological research.

[2] See Kupchik et al. 2021, especially "Biography ..." IX-XIV, "List of Publications..." XV-XXIX, Tabula Gratulatoria XXX-XXXIII, "Acknowledgements" XXXIV, and "Introduction" (by Kupchik, 1(-2)).

[3] *Italics* afm for the purposes of this review; Vovin's 2020 grammar expands over 1339 pp.

[4] Actually two (Umetomo Saeki 1933 and Noriyuki Shirafuji 1987), the former republished in "several new and enlarged and improved editions" which "can be used as a very good introduction to Old Japanese, but [...] it still remains too sketchy in many details" (*ib.*). The only grammar, described as "the main and the most detailed" but "accessible only to the people who have mastered Classical Japanese" (*ib.*) was Yoshio

---

conclude his short survey with the sentence "There is only one general grammatical sketch of Old Japanese written in a Western language (Syromyatnikov 1972)" (*ib.*, 21)[5].

The "only" Western-language sketch referred to above does but mention dialect differentiation: "Records of the 8th century do not reflect all the dialects, but mainly the Central Western and, to some extent, the Eastern [...]. In the present work, Eastern dialectal features are rendered sporadically because of lack of space" (Syromiatnikov 1981:13[6]). "Mention[ing] just the *most fundamental ones*" among writings with Eastern Old Japanese in the focus of their authors' attention, Vovin & Ishisaki-Vovin (2022: IX, also 483-4, 486-7) specifically refer to four such works, three by "Japanese" scholars (Yoshisuke Fukuda 1965, Tadao Hōjō 1966, and Yoshiharu Mizushima 1984) and only one by "Western" scholar: John Kupchik (2011).

The latter is a really fundamental and monumental, also in size (over 1050 pp.), PhD dissertation, considered "unpublished"[7] in Vovin & Ishisaki-Vovin 2022: 486 despite being globally e-accessible. The title announces "a grammar of *Eastern Old Japanese dialects* [plural]". Interestingly, all the three other works by Japanese authors referred to by the Vovins as "fundamental" include the word *Azuma* in their titles[8]. The lexeme itself is quite tricky. Generally, it is a proper noun~name, personal and geographic~topographical, and historical. As an *anthroponym*, it can function as a family name written in a number of "untypical~unusual ways" (the most frequent being 東, but also 吾妻, あずま, 吾嬬, 阿妻, 我妻, 阿室, 安部間, 阿部摩, 四阿), or its constituent (like e.g. in *Azumai* 東井, *Azumaji* 吾妻路~東路~東道~東地, *Azumaya* 東家~東谷~東屋, *Azumana-*

Yamada's "History of the Nara period grammar" of 1913 (with "several enlarged and corrected editions and many reprints"; Vovin also warns that "Yamada's grammar does not really differentiate between Western and Eastern Old Japanese, treating them as if they were data from the same language", although Yamada was aware of the grammatical differences (*ib.* also p. 1242).

[5] A very competent opinion: the "Western language" in this case is Russian, Vovin's native tongue. It was published in English translation in 1981 (in Bentley's 2001: 2 opinion, it was " rather poorly translated from the original Russian").

[6] In the original (1972: 9): "В памятниках VIII в. отражены не все диалекты, а в основном централь-но-западный и частично восточный [...]. В данном очерке по недостатку места восточные диалектизмы приводятся лишь спорадически".

[7] In this writer's opinion, "traditionally": seemingly, for a larger part of scholars more advanced in years who were professionally active in the pre-internet epoch treating PhD and other degree-dissertations as "unpublished" was obvious, hence using the attribute in reference lists was mechanical. This writer found it not only in the book just quoted. Vovin's name appears on the title page of Kupchik 2011 as that of <Chairperson of the Dissertation Committee at Hawai'i>, cf. as well "Acknowledgements" (*ib.*, p. iii).

[8] Fukuda 1965: *Nara jidai Azuma hōgen-no kenkyū* [a study of the Eastern Japanese dialects in the Nara period]; Hōjō 1966: *Jōdai Azuma hōgen-no kenkyū* [a study of the Eastern Japanese dialects]; Mizushima 1984: *Man'yōshū Azuma uta-no kokugogakuteki kenkyū* [a linguistic study of the *Man'yōshū*'s eastern poems]; all readings and translations, with minor cosmetics, taken from Vovin & Ishisaki-Vovin who listed also three other works by Mizushima with similar titles including <*Azuma*>. Evidently the title of the same 1966 work by Hojo (北条忠雄) – 上代**東国**方言の研究 in the original orthography - was in Kupchik 2011: transliterated as <*Jōdai tōgoku hōgen-no kenkyū*>, although in the case of Fukuda 1964 (奈良時代**東国**方言の研究) and Mizushima 1984 (萬葉集**東**歌の国語学的研究) the transliterations of both Vovins 2022: 483 and 487 and Kupchik 2011: 1014 and 1016 are the same. *Nota bene*, for Hojo 1966 both Kupchik 2011 and 2023 indicate different Publisher (Maruzen) than Vovin 2020 and Vovins 2022 (Nihon gakujutsu shinkōkai – the same as indicated on the book in this writer's eyeshot).

da 東洋), or also as a (masculine) given name (e.g. 東, 吾妻, 雷, or アズマ in comics and for fictitious characters in other arts; see e.g. Nichigai 1990: vol. "family names" 11, 260, 374, vol. "given names" 7-9, 402-3, 550). As a *toponym* it appears (or appeared in the past until recent administrative changes) in names of mountains and volcanoes (e.g. Azumayasan四阿山2384 (Gumma), *Azuma kofuji*吾妻小富士1707 (Fukushima), Azumayasan 四阿屋山 1387 (Nagano), Azumaneyama東根山928 (Iwate), Azumayasan四阿屋山771 (Saitama), Azumadake 東岳684 (Aomori), Azumanesan 東峰山619 (Iwate), Azumamoriyama 東森山153 (Miyagi), but also the whole volcanic *Azumayama* 吾妻山Mountain Range (*Azuma kazan*吾妻火山 ~ *Azuma rempō* 吾妻連峰)⁹, villages東村and towns 東町(e.g. *Azuma-machi* in Inashiki, Ibaraki, till 2003, also *Azuma-chō* in Izumi, Kagoshima, till 2006) throughout Japan and as a constituent of endemic plant names (like *azumagiku* 東菊 [~*ezogiku* 蝦夷菊] *Aster dubius*; *azumanezasa* 東根笹*Pleioblastus chino*; *azuma ichige* 東一華 *Anemone raddeana*; *azuma shakunage* 東シャクナゲ*Rhododendron degronianum*), names of hotels and resorts (like e.g. *Sado Resort Hotel Azuma* on Sado Island or capsule hotel *Ryoma Higashi Azuma* and several others in Tokyo, not to speak of objects like *Hotel Azzun Orient Spa & Wellness* in Mazury Lakeland of north-eastern Poland with Website also in Japanese), boats, warships, even pieces of women's wear (like footwear *azumageta* 東下駄 or overcoat *azuma kōto* 東コート). Close relations between anthroponyms and toponyms¹⁰ have been observed in very many (perhaps most) among the world's languages (i.e., probably still less than 20% of them adequately studied, described, and known), hence all the lexemes quoted above in all probability can be treated as hyponyms to be reducible to (or to have their source in) their hyperonym ("common semantic denominator") being the historical toponym *Azuma* defined by the immortal Edmond Papinot in his *Historical and Geographical Dictionary of Japan* as "the name formerly given to the 15 eastern and northern provinces of Honshū" ¹¹. Great Japanese-English *Kenkyusha* dictionaries explain the entry word "Azuma" as "east; East [Eastern] Japan; the eastern provinces; the East land" (4ᵗʰ Edition 1974) and "the Kantō region [but in Japanese as "the old name of the Kantō region": 関東地方の古称 *Kantō chihō-no koshō*]; Eastern [East] Japan; eastern provinces; *Azuma-no kuni* 東の国 the lands east of the Osaka barrier 逢坂の関より東 *Ōsaka-no seki-yori higashi*; the Kantō region; Eastern [East] Japan; the eastern provinces". Great dictionaries record also other historical compounds and phrases of interest (not only to linguists) like highly contrasting *azuma-otoko* (東男) and *azuma-ebisu* (東えびす), the former positively underlining

---

⁹ Examples in part quoted from Tokuhisa et al. 2011: 26-7; numbers indicate elevation above sea level in meters (m.a.s.l.).

¹⁰ Personal names are often associated with names of places where their bearers come from and place names are often associated with the names of their discoverers, founders, owners, etc.

¹¹ P. 43 in the *First* [Charles E.] *Tuttle* [Company] *edition, Eighth printing* 1984 used. Papinot (1860-1942) related it "etymologically" with the legendary exclamation "*Azuma wa ya!* [吾妻はや <吾妻] 'ah, ma femme!'". His dictionary was first published in 1899 in French as *Dictionnaire japonais-français des noms principaux de l'histoire et de la géographie de Japon* and its English version released in 1906 is in consecutive editions and reprints on sale till these days.

masculinity (~machismo), the latter, clearly discriminatory, derogatively implying primitivism, cultural inferiority, barbarousness, hence contempt for persons spoken about[12].

In a very short "Preface" (pp. V-VI; one would definitely welcome its author to be much less economical in his words) to his *Grammar* presented here, and in the initial part of its "Introduction" (1-2), Kupchik defined his understanding of the term *Azuma*, or rather its understanding for the purposes of compiling his grammar (initially planned to be "a book about the phonology and phonetics", only "later deciding" that it would be better to expand it into a full grammar" (V) – general linguists will be long blessing and praising him for such decision). In his words, the modifier *Azuma* in the glottonym <Azuma Old Japanese> is a toponym: "It should be noted from the outset that" he uses the glottonym (abbreviated to) ""AOJ" as a purely areal term" (1). This reviewer, however, eagerly looks at it as if at a new internationally acceptable language name, like e.g. **Azuman* (an extinct member of *Japonic languages*), on the horizon.

The author of the book returns to the term in the first sentences of the "Introduction": "The Azuma Old Japanese (...AOJ) dialects were spoken in Japan during the Nara period (8th c. CE) in the eastern region called Azuma that stretched from present-day Shizuoka and Nagano, east to Ibaraki, and all areas between them extending southward to the Pacific. The large northeastern area [...] that covered present-day Fukushima, Miyagi, Iwate, and Aomori" Prefectures "was also included" [...] (p. 1).

The date (1999)/2000 marked not only the turn of the century, and the millennium, but also the turn of the epoch in the history of Western research in/on the oldest attested~recorded stage of what gradually developed into the <modern Japanese language>. The 170 pages of ("only one") Syromyatnikov 1972 (140 pages in the 1981 English

---

[12] Both Kenkyusha 4th and 5th editions translate the former as "a man from East(ern) Japan" ("regarded as more manly than those from the area around Kyoto", 5th, p. 45) and illustrate it with the phrase *azuma-otoko-ni kyōonna*, translating it "the best men are found in Edo and the best women in Kyoto" and interpreting it "Go to East Japan for manly (masculine) qualities; go to Kyoto for womanly (feminine) beauty" (4th) and "Eastern Japan for manliness; Kyoto for feminine charm" (5th). The latter is translated in Kenkyusha 4th as "savages in the eastern provinces" and the *ebisu* component (noted with the character 夷) in a separate entry is rendered as "barbarian; a savage"; in Kenkyusha 5th it seems to be absent but both editions include the entry *tōi* (東夷) found also in Kenkyusha 2nd edition (here illustrated with two other very interesting historical terms: *tōi seibatsu* 東夷征伐 "the subjugation of the eastern barbarians" and *tōi sei shōgun* 東夷征将軍 "a commander of an expedition against the eastern barbarians; a military commissioner of the barbarian east") and in many other monolingual (cf. e.g., Kodansha's 1989:39 elegant explanation of the entry *Azuma ebisu* as 京都の人が東国の武士をさげすんでいった語 'a term used by Kyotoites to express looking down upon ~ spurning the warriors of the eastern provinces') and bilingual (including Nelson and New Nelson, "a must" on the desk of almost every student of Japanese) dictionaries. Both the word *ebisu* and the character 夷 (read *ebisu* or *i*) were in the past predominantly used as ethnonyms for the 'Ainu'. These, longish and possibly perceived as superfluous, lecture-like elucubrations in this footnote resulted from this writer's astonishing observation that the historical toponym *Azuma* and its graphic notation as well as terms associated with it were not known even to many graduates in Japanese studies; the astonishment vanished immediately upon finding the absence of not only the appropriate entry in the first volume of the eight-vols. *Kodansha Encyclopedia of Japan* (and its first *Supplement* (1986), too) but possibly even a single occurrence of the term throughout the entire 1983 edition as an inspection of its *Index* would suggest (while the entry "Satsuma Province" *Satsuma-no kuni*, considered in Japanese literature as opposite (at least "phonetically": -*tsuma*~-*zuma* implied 'end, edge, verge, fringe, margin, rim, brim, brink, tip, side, shore, extremity, border, frontier, peripheries') to "Azuma provinces" *Azuma-no kuni*, does have its place in vol. 7). Flash haphazard checking in a number of "histories of Japan", "dictionaries of Japanese history" included, concluded with the same effect.

translation) mentioned above were replaced by the whole library of bulky publications of fundamental significance for the discipline (like e.g. Bentley 2001, Miyake 2003, Kupchik 2011, Vovin 2009-20, ²2020 (¹2005/I-2009/II), Vovin & Ishisaki-Vovin 2022, Kupchik 2023 focused on in the present text, Frellesvig & Kinsui 2024) as well as an eruption and proliferation of minor supportive contributions (e.g. Kupchik & Alsonso & Miyake 2011:VVIII-XXIV, 5-64; resp. chapters in Frellesvig & Kinsui 2024) and numerous journal articles (e.g., listed in bibliographies attached in the works just mentioned). Research in the field was so intensive that "revised, updated, enlarged", etc., editions proved necessary within a decade or two (cf. fn. 13).

Kupchik 2023 belongs also to this latter category, being (as expected) a "rewritten" Kupchik 2011. "Looking back on" his "dissertation several years later", its author "realized it needed many significant revisions, rewrites, and additional analyses" (V). Such a sequence of events, approaches, and receptions of one's own work should be considered normal and typical in academic biographies and dissertations should be first preprinted for bureaucratic (officious) procedures and a "several years" buffer period before their final much more mature on many obvious reasons) publication proves always beneficial. The author's "looking back" turned out to be very critical as he openly admits: "There were many typographical errors, topics with insufficient discussion, and my *views on certain aspects the data had evolved or changed completely over time*". Rewriting involved additions of "a *significant amount of new text*" and "a few new chapters" but also replacement of "more than 400 pages [cf. the total number of pages mentioned above] of phonological comparison charts" with their summaries[13], simultaneously reminding readers particularly interested in details about their availability in Kupchik 2011 (*ib.*, *italics* afm.).

The front matter of the book introduced here expands over 34 pages ([I]-XXXIV) and includes *i.a.* "Preface" (cf. above), two tables of contents (the first, IX-XI, standard size, and the other "Extended table of contents", XIII-XXVIII), "Glossing [and] translation conventions" (XXIX), and "Abbreviations" ("Grammatical Terms", XXXI-XXXII; "Azuma Old Japanese Provinces /Topolects", XXXII-XXXIII; "Language varieties", "Texts", "Primary Manuscripts", XXXIII; "Other" (three items), XXXIV). The idea to introduce the two tables of contents is an interesting logistic proposal welcomed and recommended for more frequent application[14] in similar publications: each of them performs a different function[15].

The core of the monograph is organized into ten units (the author calls them "chapters"[16]), the first of them being "Introduction" (1-21) which informs on grammatical and

---

[13] This reviewer shares in full Kupchik's opinion that it "would be easier to read and more useful for researchers" (especially those not intending to specialize in studying the history of the Japanese language to whom the present review is addressed in the first place).

[14] But not overuse: probably in a majority of similar cases a good extensive index (or indices) surely is a better, more user-friendly, solution.

[15] And both can be irritating, when only one of them is present and index is meager: while one is too laconic, not informative enough, the other is too time-consuming in use and often purposelessly increases redundancy. See further in this text mentioning the full title of the "Index".

[16] This reviewer is perhaps too old to get accustomed to treat introductions (also prefaces, conclusions, appendices, etc.), as <chapters> (in his world order, except for <preface>, they follow the <introduction>).

typological features of the language described as well as on its place on the linguistic map and among other -lects of the eighth century Japan, but also on the state of the art in *Azuman* philological and linguistic research. It also defines the "structure and aims" of the book and, since it is to be a grammar of a language (one of the -lects) in use more than a millennium ago, long since extinct but/and preserved in poetry only, briefly explains or informs on sources, manuscripts, versification, metrics, writing system(s) involved, phonetic and phonemic reconstructions, orthography and transliterations, to conclude with specifying the author's goal – the key terms used are <descriptive grammar> and <reference grammar> (the latter being used in relation to chapters 3 to 10 – "the second section of the book"; 21) – *exactly* what *general* linguistics and linguists are in need of as far as Old Japanese is concerned.

[Chapter] 2 (23-104) constitutes a study of post-Proto-Japanese period sound changes involving reconstructions of several proto-lects (like Proto-Japanese, Proto-Eastern Old Japanese, Azuma Old Japanese topolects) and "numerous protoforms" to propose "a new subgrouping of the Old Japanese dialects" and "a detailed grouping of Eastern Old Japanese topolects. This chapter has many changes and revisions to the reconstruction and analysis presented in Kupchik" 2011 (23). [Chapter] 3 (105-17) entitled "Vowel elision, *rendaku*[17], and assimilations" is also concentrating on phonemics, but the description is synchronic and focusing on Azuman.

[Chapter] 4 (119-42) "Lexicon" devotes about one page (119-20) to the native (Japonic stock) vocabulary, two and half pages (120-2) to Korean, five lines (140) to Chinese, a little over one page and half (141-2) to Austronesian, and eighteen pages (123-40)[18] to Ainu loanwords[19].

What follows is exactly what one expect to find in the first place in any grammar of any language one reaches for to consult. [Chapter] 5 treats "Nominals" ((143-243), subgrouped into nouns (143-209), pronouns (209-37) and number names (237-43, with clossifiers); [chapter] 6 (245-73) describes uninflective adjectives (one example of a reduplicated adjective found, 251) and conjugable "adjectival" verbs (no tense and aspect); [chapter] 7 (277-426), "Verbs" constitutes the longest unit in the book, and no wonder: its first sentence confirms the expected: the verb indeed is "the most morphologically complex" lexical category in the language described – with grammatical categories of *i.a.* negation, aspect, tense, mood[20], voice, retrospection, honorification, rich inventory of categorial markers (*i.a.* affixes (292-383), auxiliaries (384-419), reduplication (425-6)). This "reference" part of the grammar ends with three short chapters on, perforce com-

---

[17] 'sequential voicing' 連濁, the voicing of initial consonants in the non-initial components of compound words (e.g. *hitobito < hito + hito, tegami < te + kami, daidokoro < dai + tokoro, benjo < ben + sho*; in Old Japanese prenasalization was additionally involved.

[18] 75% of the chapter text – the foremost important argument for this reviewer to secure a permanent place for Kupchik's *Azuman* grammar on one of the more prominently exposed (still not clear, be it "Ainu" or "Japanese") shelves of his library.

[19] In the case of Ainu, it is a "critical assessment of [Vovin's] proposals" (123) categorized into toponymic, lexical, and morphological borrowings, and into "plausible (accepted)" and "problematic (rejected)"

[20] "Arguably the most notable characteristic of [Azuman] verbs is the rich system of mood markers" (277).

mented lists of: 8 "Adverbs" (427-34, including adverbial constructions), 9 "Conjunctions" (435-40, and "Particles" (441[21]-82).

No special unit for syntax in the book but all relevant information (like primary and secondary word order, verb serialization, reduplications, particle placement) can easily be found in the units presented to readers above.

The final units of the book are: an appendix ("A classification of poems in [*Man'yōshū* poetic anthology] Book 14 based on linguistic features" with 379 footnote comments, 483-509), list of "References" (511-17), and "Index (supplement to the Extended Table of Contents)" (519-22).

To be sure, much more time than that passed since the publication date of the book under review[22] and much more experience with staying and getting familiarized with it, consulting it, using for purposes now not to be named, is needed to more adequately and more objectively evaluate and appreciate its usefulness (especially, if its user does not aspire to get deeply involved either in the history of the Japanese language or, say, translating the *Man'yōshū* poetry). Nevertheless, this reviewer's early impressions from closer several hours long inspection of the volume are very positive and promising, the structure of argumentation makes the reading pleasant, and the transparent (often tabularized) presentation of the Azuman linguistic material strengthens the user's trust in its reliability. Definitely, with the release of this Kupchik's grammar we now have at our disposal what we never had before. It elegantly (also visually) complements the recent (21st-century) "whole Old Japanese library" mentioned above: phonetic and phonemic reconstructions, three grammars (Early Old Japanese prose, Western Old Japanese, Eastern Old Japanese), and two dictionaries (Eastern Old Japanese and Ancient Japanese phonograms)[23].

What general linguists are still lacking, however, are – suitable for them – dictionaries of earlier stages of Japanese[24].

References

Bentley, John R. 2001. *A descriptive grammar of early Old Japanese prose* (Brill's Japanese Studies Library 15). Leiden–Boston–Köln: E. J. Brill.
Bentley, John R. 2016. *ABC dictionary of Ancient Japanese phonograms*. Honolulu: University of Hawai'i Press.

[21] One more quotation: "Particles are a fundamental aspect of the grammar of the [Azuma Old Japanese] dialects. [They] can be separated into seven classes: focus..., emphatic..., question..., restrictive..., desiderative..., ironic..., and quotative particles" (441).

[22] Oct. 24, 2023; this text typed May 2024.

[23] We avoid here discussing or listing electronic-only tools, seemingly quite numerous and seemingly not attractive to "non-Japanologist" linguists, like < Bjarke Frellesvig & Stephen Wright Horn et al. (eds.) 2023. *Oxford-NINJAL Corpus of Old Japanese* available at: http://oncoj.ninjal.ac.jp/> (probably the best source and tool in the making for Japanologist linguists).

[24] Like Old Japanese-to- [Western, e.g. English], but also... Classical Japanese (文語~古文)-to [Western], arranged user-friendly, perhaps with entries arranged like in e.g. standard *Kenkyusha* dictionaries of Contemporary Japanese (transliteration – original orthography – equivalents in translation, preferably with examples and comments) but offering material of e.g. popular and indispensable in Japan古語辞典*kogo jiten* 'dictionaries of Classical Japanese'. The lack of such dictionaries seems unbelievable.

Frellesvig, Bjärke & Satoshi Kinsui (eds.). 2024. *Handbook of historical Japanese linguistics.* HJLL, vol. 1. De Gruyter Mouton.

Fukuda Yoshisuke. 1965. *Nara jidai Azuma hōgen-no kenkyū* [studies on Nara period Eastern Japanese dialects].Tōkyō: Kazama Shoin. 福田良輔. 1965。奈良時代東国方言の研究。東京: 風間書房。

Hōjō Tadao 1966. *Jōdai Azuma hōgen-no kenkyū* [studies on Eastern Old Japanese dialects]. Tōkyō: Nihon Gakujutsu Shinkōkai. 北条忠雄 1966。上代東国方言の研究。 東京: 日本学術振興会。

Kenkyusha – *Kenkyusha's new Japanese-English dictionary* editions used: Takenobu Yoshitaro (ed.) (²1931, 91st printing) 1941; Koh Masuda (ed.) 1974. 4th edn; Watanabe Toshiro [&] Edmund R. Skrzypczak [&] Paul Snowden (eds.) 2003. 5th edn. Tokyo: Kenkyusha.

Kodansha – Umesao Tadao [&] Kindaichi Haruhiko [&] Sakakura Atsuyoshi [&] Hinohara Shigeaki (eds.) 1989 (second printing 1992 used). *Kōdansha karāhan Nihongo daijiten* [great dictionary of the Japanese language with color illustrations]. Tōkyō: Kōdansha. 梅棹忠夫·金田一春彦·阪倉篤義·日野原重明監修 1992。講談社カラー版日本語大辞典。第二刷発行。東京: 講談社。

Kupchik, John E. 2011. *A grammar of the Eastern Old Japanese dialects*. [Honolulu: University of Hawai'i at Mānoa Department of Linguistics]. (Doctoral dissertation.)

Kupchik, John & José Andrés Alonso de la Fuente & Marc Hideo Miyake (eds.). 2021. *Studies in Asian historical linguistics, philology and beyond: Festschrift presented to Alexander V. Vovin in honor of his 60th birthday*. Leiden–Boston: Brill.

Majewicz, Alfred F. 2022. Review of Vovin & Ishisaki-Vovin 2022. *Rocznik Orientalistyczny* 75(2). 155-9.

Miyake, Marc Hideo. 2003. *Old Japanese: A phonetic reconstruction*. London: RoutledgeCurzon.

Mizushima Yoshiharu. 1984. *Man'yōshū Azuma uta-no klokugogakuteki kenkyū* [studies on *Man'yōshū* eastern poems from the stand of Japanese (national) linguistics]. Tōkyō: Kasama Shoin. 水島義治 1984。萬葉集東歌の国語学的研究。東京:笠間書房。

Nelson, Andrew N. (²1974, 17th printing) 1984. *The modern reader's Japanese-English character dictionary. Second Revised Edition.*

Nichigai Associates (eds.). 1990. *Nihon seimei yomifuri jiten* [in two volumes]. *Sei-no bu // Japanese family names in Chinese characters. Mei-no bu // Japanese first names in Chinese characters: A guide to their readings*. Tōkyō: Nichigai Associates, Inc. 日外アソシエーツ編集·発行 1990。日本姓名よみふり辞典。姓の部。 名の部。東京: 日外アソシエーツ株式会社。

Saeki Umetomo. 1933. *Jōko-no kokugo* [cf. fn. 4]. Tōkyō: Meiji Shoin. 佐伯梅友著1933。上古の国語。國語科學講座5。國語史學。東京:明治書院。

Shirafuji Noriyuki. 1987. *Nara jidai-no kokugo* [cf. fn. 4]. Tōkyō: Tōkyōdō. 白藤禮幸1987。奈良時代の国語。東京堂出版。

Syromiatnikov, N[ikolay] A[leksandrovich]. 1981. *The ancient Japanese language*. Moscow: Nauka Publishing House. Translation of: Н[иколай] А[лександрович] Сыромятников 1972. *Древнеяпонский язык* [Old Japanese langugae]. Москва: Наука.

The New Nelson Japanese-English character dictionary based on the classic edition by Andrew N. Nelson completely revised by John H. Haig. 1997. Rutland, Vermont & Tokyo: Charles E. Tuttle Co.

Tokuhisa Tamao [&] Ishii Mitsuzū [&] Takeuchi Tadashi (eds.). 2011. *Sanseidō Nihon sammei~yamana jiten, kaiteiban* [Sanseido encyclopedic dictionary of Japanese mountain names, revised edition]. Tōkyō: Sanseidō. 徳久球雄 · 石井光造 · 武内正編集2011。三省堂日本山名事典。改訂版。東京: 三省堂。

Vovin, Alexander. 2020. *A descriptive and comparative grammar of Western Old Japanese: Revised, updated and enlarged 2nd edition*, vols. 1-2 (Handbook of Oriental Studies // Handbuch der Orientalistik. Section Five Japan 16/1-16/2). Leiden–Boston: Brill. Pp. XXXII + 1-438 and XIX + 439-1278, respectively.

Vovin, Alexander (transl. ed.). 2009-20. *Man'yōshū (Book* xx*). A new English translation containing the original text, Kana transliteration, Romanization, glossing and commentary* . Books: *1*(2017), *2*(2020), *5*(2011), *10* (to be released posthumously), *14*(2012), *15*(2009), *16*(2020), *17*(2015), *18*(2016), *19*(2018), *20*(2013). Brill & Globar Oriental [a Brill's imprint].

Vovin, Alexander and Sambi Ishisaki-Vovin. 2022. *The Eastern Old Japanese corpus and dictionary* (Handbook of Oriental Studies // Handbuch der Orientalistik. Section Five Japan. Volume 17). Leiden–Boston: Brill.

Yamada Yoshio. 1913. *Nara-chō bumpō shi* [cf. fn. 4]. Tōkyō: Hōbunkan. 山田 孝雄 1913。奈良朝文法史。東京: 寶文館。(1954 revised edition改版)。