

## Lexicostatistical Studies in East Sudanic I: On the genetic unity of Nubian-Nara-Tama

In this paper, I present a detailed lexicostatistical survey of the reconstructed 50-item word-lists (the “more stable” half of the classic Swadesh list) for three language groups of North-east Africa — Nubian, Nara, and Tama, commonly ascribed to the East Sudanic family and often described in related literature as forming a specifically tight-knit node within that taxon. The survey shows that both the number and the nature of direct lexicostatistical matches between these three groups is plausibly interpretable as decisive evidence for genetic relationship, adding one more formal confirmation to the evidence previously assembled by J. Greenberg, M. L. Bender, Claude Rilly and other scholars. Glottochronological interpretation of the evidence, however, indicates that Nubian-Nara-Tama should be dated to at least the 5th millennium BC, which makes it older than Indo-European and presumably very hard to reconstruct in sufficient detail. The paper itself is the first in a series of planned publications that will explore the East Sudanic hypothesis from a combined lexicostatistical and etymological perspective.

*Keywords:* Nilo-Saharan languages, East Sudanic languages, Nubian languages, Tama languages, African historical linguistics.

### General introduction

Of the three macrofamilies that Joseph Greenberg had delineated in his seminal works on African language classification (most importantly Greenberg 1966<sup>1</sup>), the “Nilo-Saharan” taxon has always shared the most vague outlines. While Greenberg’s “Niger-Kordofanian” languages are informally understood as “the ones with the complex noun class systems” (subsequently, the few subgroups that violate this feature, such as Mande, are sometimes viewed with suspicion even by supporters of the Niger-Kordofanian hypothesis<sup>2</sup>), and Greenberg’s “Khoisan” is just as informally understood as “the click family”, there are no such definitive features to characterize all, or even the majority of the language groups that, according to Greenberg, constitute the Nilo-Saharan macrofamily: the hypothesis is based on numerous, if not properly systematized, lexical and grammatical resemblances rather than any structural homologies.

This fact in itself is not necessarily problematic for historical linguists, since it is commonly accepted, and has frequently been pointed out by Greenberg himself, that genetic relationship is not to be established based on typological features of languages, easily open to areal influence (cf. the spread of “Khoisan” click phonemes to neighboring Southern Bantu languages), but should always be defined primarily by the presence of important homologies

<sup>1</sup> The fourth macrofamily — Afro-Asiatic, formerly known as Hamito-Semitic — was already more or less securely recognized as a genetic unity long before Greenberg’s works, and may be kept out of any general discussion on the overall quality of Greenberg’s methods and arguments.

<sup>2</sup> Blench (2011) presents a seemingly strong case for the innovative nature of nominal class markers in the bulk of NK, but this view has not yet gained extensive support from specialists.

in the phonetic structures of lexical and grammatical morphemes bearing identical or similar meanings. To that end, Greenberg's argumentation in favor of his macrofamilies always consists of comparative lists and tables of such morphemes. Nevertheless, typological considerations still continue to play an important part in the general acceptance of macrofamily hypotheses — if anything, they offer intuitive support in situations where form-based arguments are either too complex or too dubious for us to quickly assimilate and evaluate. Since the primary methodology behind Greenberg's macrofamilies has been that of “mass comparison”, commonly criticized by linguists as a procedure that is unable to properly separate genuine traces of genetic relationship from either areal contacts or chance similarities, it is not surprising that his comparative lists of words and morphemes do not seriously impress modern specialists, whereas such features as the presence of click phonemes or noun class markers do — at the very least, such structural homologies cannot be easily explained away as accidental resemblances.

In this type of situation, linguists who properly dedicate themselves to the construction of an optimal scenario of genetic relationship in a particular linguistic area should find it of essential importance to define specific sets of “genetic markers” (a term that seems quite naturally borrowable from molecular biology) that concisely characterize the postulated taxon and distinguish it from its neighbors. Roughly speaking, such markers should:

- (a) constitute either grammatical morphemes or lexical roots that belong to the basic (i.e. generally more resistant to diachronic change) layer of language;
- (b) be reconstructible for *all* or *most* of the proposed subbranches of the taxon (at the very least, be reliably reconstructible in its most distant branches, to assure their protolanguage status);
- (c) respect the general laws of phonetic change, suggested for the taxon, or, if the taxon is a high-level one, at least yield reflexes in daughter branches that could be deemed “phonetically compatible”, i.e. explainable through typologically and historically realistic scenarios of phonetic change<sup>3</sup>;
- (d) demonstrate either the *exact same* meaning in all or most of the daughter branches, or display minimal semantic variety, confined to diachronically and synchronically frequent types of semantic change or polysemy found in the world's languages (such as ‘eye : see’, ‘black : dark’, ‘know : hear’, etc.)<sup>4</sup>;
- (e) preferably, at least some of them should be *exclusively* representative of the suggested taxon, in that it could be at least approximately demonstrated that they are reconstructible in that particular form and meaning for the proto-language of that particular taxon and no other.

For linguistic taxa that have diverged within the last five or six thousand years and whose linguistic history has been reasonably well studied, due to an abundance of both primary data and analytical research, the presence of such genetic markers is an obvious fact — a lexical root such as, e.g., Proto-Indo-European \**okʷ-* ‘eye’ satisfies all of the listed conditions. For speculative linguistic “macrofamilies” whose hypothetical age goes far beyond the specified chronological range, producing such markers is a highly complex challenge, since the prob-

---

<sup>3</sup> A detailed explanation of the idea of “phonetic compatibility” and its difference from both the weaker criterion of “phonetic similarity” and the stronger criterion of “phonetic correspondence” may be found in Starostin 2013: 57–64.

<sup>4</sup> Although, as of now, there is still no single definitive list of such polysemies that would be both sufficiently comprehensive and obtained through a formal methodology, progress is slowly being made with such works as Youn et al. 2016. As far as basic lexicon is concerned, careful fixation of attested polysemies is conducted by contributors to the Global Lexicostatistical Database project, which allows to perform rough statistical estimates of what may count for a “trivial” polysemy or semantic shift.

ability of their successful recovery decreases with each added millennium. Nevertheless, even a highly limited set may be convincing if it can be shown to have been arrived at without any distortions of available evidence or violations of known tendencies of language change through idiosyncratic assumptions.

In the case of Nilo-Saharan, the proper search for such “genetic markers” was originally launched by M. Lionel Bender, whose sets of “excellent”, “good”, and “fair” isoglosses (Bender 1997: 77–105), assembled in favor of the hypothesis, satisfy some of the above-listed criteria. However, even some of his “excellent” isoglosses play quite loosely with semantics (e. g. such connections as ‘elbow/claw/foot’ or ‘horn/bone/rib’ are quite suspicious) and remain uninterpretable in terms of reasonable historical scenarios of semantic change; numerous phonetic deviations are recorded without any attempts at constructive explanations; and, perhaps most importantly, a huge number of comparanda are not shown to be reconstructible for the required intermediate levels of comparison, which means that they have been too quickly transferred to a deeper level of comparison without proper completion of the preceding stage of analysis — and, consequently, without a reliable “safety net” against accidental resemblances.

The late Lionel Bender himself may have been well aware of these limitations of his own research; in any case, it is somewhat instructive that, instead of expanding his relatively short overview monograph on Nilo-Saharan (Bender 1997) to the size of an etymological dictionary (such as the huge, but ultimately unconvincing volume by Christopher Ehret (2001)), he preferred to follow it up with an equally short comparative treatise on East Sudanic (Bender 2005) — a pioneering study, focusing on one of the largest sub-taxa originally defined by Greenberg within Nilo-Saharan.

The natural implication behind Bender’s East Sudanic book is that, without a proper understanding of what exactly is “East Sudanic”, we cannot gain any understanding of what exactly could be “Nilo-Saharan”. Ironically, in his introduction to the book, Bender mentions having been unable to establish an “East Sudanic Working Group”, since “the main problem seems to be that no one is willing to go beyond a narrower focus on sub-families” (p. vi). Indeed, genealogical nodes like East Sudanic find themselves in double trouble: the proverbial “splitters” (or simply specialists with a narrow focus) are not interested in working on them because the explored genetic connections are seen as too deep and complicated to recover, whereas the proverbial “lumpers” (linguists with a pronounced interest in macro-comparative studies) view them, at best, as quick stepping stones, postulated mainly for the sake of classificatory convenience, then more or less forgotten as the interest rapidly shifts to highest-level taxa.

The only work other than Bender’s all-too-brief monograph that actually tries to tackle East Sudanic on a serious basis seems to be Rilly 2009, which includes a very thorough comparative analysis of the phonological systems and lexica of those branches that, according to the author, constitute the “Northern” division of this family, including Nubian, Tama, Nara, and Nyimang. However, even in Rilly’s book, the arguments in favor of East Sudanic are not really assigned any stand-alone value; rather, they are considered significant inasmuch as they help determine the genetic affiliation of the Meroitic language, which, based on scarce evidence of often dubious quality, Rilly seeks to relate to “Northeast Sudanic” (including Nubian, which seems to have the strongest links with Meroitic, although it still remains unclear whether most of them are of a genetic or areal nature). Furthermore, dealing with but one branch of East Sudanic is certainly not the same thing as trying to evaluate the validity of the entire family.

It was mostly these considerations that eventually led to a general lexicostatistics-based survey of possible genetic connections between the various groups of languages that constitute Greenberg’s “Nilo-Saharan”, in which the East Sudanic hypothesis was tested first — without taking into account any higher level connections. The test, carried out as part of a

large ongoing project on the general classification of African languages, followed a standardized methodology that had already been tried out on the so-called “Khoisan” languages, yielding results that seem to be largely consistent with current mainstream views on their classification (Starostin 2013). The main stages of this procedure may be briefly summarized as follows.

1. Define the primary constituents of the analysis. These are identified as relatively small language groupings whose genetic reality is beyond reasonable doubt and commonly accepted by all specialists — e.g., Nubian, Tama, Daju, Kuliak, etc.; all the languages within each such group share numerous cognates easily linked together with sound laws, as well as robust grammatical isoglosses, indicating a relatively recent split from a common ancestor (not to exceed 2,000–3,000 years based on any available historical, archaeological, and lexicostatistical estimates).

2. Assemble and check complete 100-item Swadesh lists for as many languages of these small groupings as possible, based on the most recent and accurate sources available. The compilation procedure closely follows the guidelines that were laid down in earlier methodological publications (Starostin 2010; Kassian et al. 2010).

3. Carry out a lexicostatistical analysis of the data in order to determine the internal classification of the groupings (most importantly, the primary splits within each of them; these results will have a direct bearing on the efficiency of point 4).

4. Reconstruct the proto-wordlist for each such grouping, based on regular etymological analysis and a complex set of criteria used to determine the “optimal” candidate for the expression of each particular Swadesh meaning in the protolanguage. Unlike wordlists for attested languages, reconstructed proto-wordlists are limited to 50 of the most generally stable Swadesh items (out of 100), since reconstruction of the second, less stable, half usually turns out to be cost-ineffective for purposes of high-level comparison and classification<sup>5</sup>. As a rule, this is the most complicated, time-consuming, and text-heavy part of the entire procedure (unless the group in question consists of several very closely related dialects that do not require detailed historical analysis).

5. Subject the reconstructed proto-wordlists to several additional stages of lexicostatistic analysis, which include running a completely automatic procedure of finding “pseudo-cognates” between reconstructions, based on the “Dolgopolsky consonantal classes” method of phonetic comparison (general description of the method and an example of its application may be found in Kassian, Zhivlov, Starostin 2015). After that, the results undergo a procedure of “manual correction” which takes into account the locally specific phonetic features of compared (proto-)languages, not recognized in the universally applicable method.

6. Compare the lexicostatistical matrices and classificatory trees generated by the “fully automated” and “manually corrected” methods and select one as the optimal choice for a working model (in most cases, this turns out to be the tree/matrix based on the “manually corrected” list of hypothetical cognates, although there may be occasional exceptions).

The current results of this procedure<sup>6</sup> are summarized in the following lexicostatistical matrix (Table 1) and phylogenetic tree (Fig. 1), both of them reflecting the “manual correction”

---

<sup>5</sup> See Starostin 2010 for additional information on how the average “stability index” for various Swadesh items was calculated and on other technical factors that have influenced the final compilation of the universally applicable 50-item list. The procedure of proto-wordlist reconstruction, illustrated by specific examples, is described in detail in Starostin 2016.

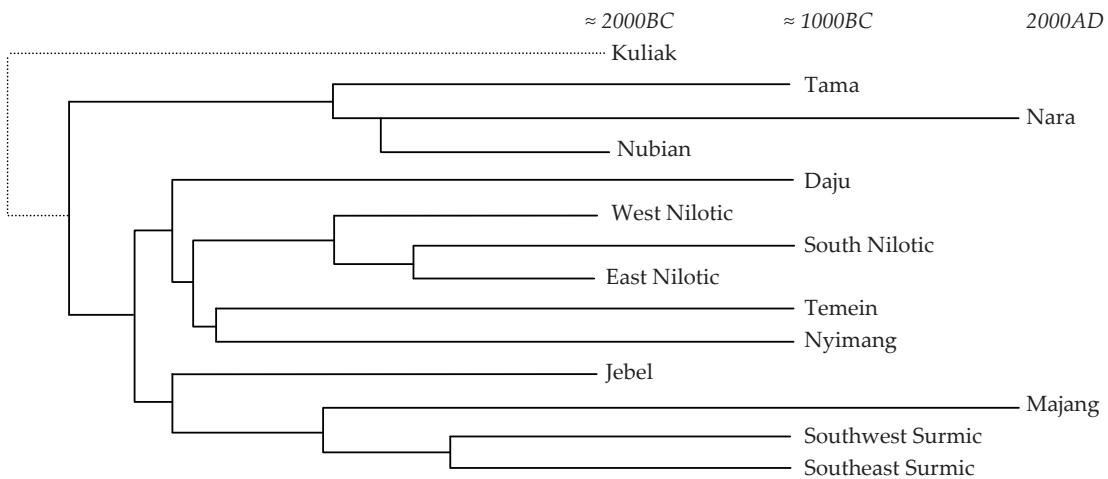
<sup>6</sup> These results differ slightly, but not crucially, from the results published earlier in Starostin 2014: 677 — an inevitable development that is due to corrections of previously produced reconstructions in the light of newly available data or occasional spotted mistakes in previous analysis. It goes without saying that these results as well are liable to future amendments, since new sources of data that allow for deeper insights become available to researchers on a steady basis.

model (which is not very different from the fully automatic model, except for the relative position of the Daju branch on the tree; this is due to certain rare types of phonetic change that took place on the way from the Proto-East Sudanic stage to Proto-Daju, some of which are quite evident even on the limited data of the 100-item wordlists).

Table 1. Lexicostatistical matrix for Greenberg's "Eastern Sudanic" (50-item wordlists).

|                  | Nara | Tama | SWS | SES | Maj. | WNil | ENil | SNil | Nyi. | Tem. | Jebel | Daju | Kul. |
|------------------|------|------|-----|-----|------|------|------|------|------|------|-------|------|------|
| Nubian           | 26%  | 20%  | 14% | 12% | 4%   | 18%  | 16%  | 20%  | 22%  | 12%  | 12%   | 4%   | 8%   |
| Nara             |      | 20%  | 10% | 10% | 8%   | 12%  | 10%  | 12%  | 12%  | 12%  | 12%   | 6%   | 4%   |
| Tama             |      |      | 6%  | 10% | 6%   | 8%   | 12%  | 16%  | 12%  | 6%   | 4%    | 6%   | 2%   |
| Southwest Surmic |      |      |     | 40% | 22%  | 16%  | 14%  | 20%  | 14%  | 14%  | 18%   | 8%   | 6%   |
| Southeast Surmic |      |      |     |     | 14%  | 20%  | 12%  | 18%  | 12%  | 14%  | 16%   | 10%  | 4%   |
| Majang           |      |      |     |     |      | 12%  | 10%  | 10%  | 10%  | 10%  | 14%   | 12%  | 2%   |
| West Nilotic     |      |      |     |     |      |      | 35%  | 18%  | 14%  | 18%  | 18%   | 16%  | 4%   |
| East Nilotic     |      |      |     |     |      |      |      | 40%  | 12%  | 15%  | 20%   | 18%  | 4%   |
| South Nilotic    |      |      |     |     |      |      |      |      | 20%  | 17%  | 14%   | 12%  | 8%   |
| Nyimang          |      |      |     |     |      |      |      |      |      | 18%  | 14%   | 12%  | 2%   |
| Temein           |      |      |     |     |      |      |      |      |      |      | 20%   | 16%  | 6%   |
| Jebel            |      |      |     |     |      |      |      |      |      |      |       | 12%  | 4%   |
| Daju             |      |      |     |     |      |      |      |      |      |      |       |      | 6%   |

Figure 1. Phylogenetic interpretation of the matrix in Fig. 1<sup>7</sup>



Both the matrix and the tree diagram suggest that, in general, Greenberg's "East Sudanic" is a viable proposition. In the majority of cases, pairwise percentages exceed 10% and sometimes rise as high as 20–25% — for a procedure that relies exclusively on phonetic similarity and inevitably omits a share of true historical cognates, this is a significant number that is very

<sup>7</sup> The tree diagram has been generated by means of the distance-based neighbor-joining method used in the StarLing software, with a glottochronological component (needed as a comparison basis for reconstructed protolanguages of varying time depths); see S. Starostin 2000 on details of the glottochronological method and Kassian 2015 for a more detailed description of the tree-building procedure. Glottochronological dates on the tree in question are only given up to the approximate time depths of all the intermediate reconstructions involved in the comparison; due to the "automated" cognate-finding procedure forming the core of the present analysis, chronological figures beyond the threshold of 3–4 thousand years will most likely be incorrect.

rarely reached under the same conditions by unrelated pairs of languages. Additionally, the results are in agreement with Bender's and Rilly's idea of a primary split into two branches (Bender's "Ek" and "En" and Rilly's "Northeast" and "Southeast" ones, respectively), with Nubian, Nara, and Tama constituting the bulk of the former; only Nyimang, which both researchers decidedly place in the "Ek/Northeast" branch, is grouped closer to Temein on the resulting tree, but this may be a phylogenetic error caused by some unrecognized convergence processes between Temein and Nyimang, an issue to be investigated later on a more thorough etymological basis.

The only glaring candidate for potential exclusion from the East Sudanic inventory is the Kuliak group: these languages consistently show around 4% to 6% resemblances with other East Sudanic branches on the 50-item wordlist — a figure that makes Kuliak as "East Sudanic" in nature as, say, the Hadza isolate (with which Kuliak languages also share 6% of superficial matches), most of which are monoconsonantal and either reflect chance similarities or, perhaps, occasional traces of much deeper relationships that are, at the present stage of analysis, indistinguishable from the former<sup>8</sup>.

Nevertheless, in order to be properly convincing, any "working model" constructed by means of preliminary lexicostatistics has to undergo further scrutiny. Even a situation where two or more languages show 20–25% of similarities on the 50-item list may theoretically be interpreted as the result of tense linguistic contact, perhaps multiplied by a few accidental resemblances. From the regular historical-comparative point of view, pure statistics is not enough: the observed and quantified similarities must satisfy our general expectations for a situation of language relationship. In particular, similarities must be organised into patterns of recurrent correspondences — a task that is often impossible to perform based on the limited material of 100, let alone 50 items, so additional material must be considered — and, if possible, additional argumentation must be presented as to why these similarities are more conveniently explained as the results of vertical rather than horizontal transmission, since regular correspondence patterns can be observed between donor and recipient languages just as frequently as between the descendants of a single protolanguage.

The chief goal of the current paper is to investigate one particular node of the preliminary lexicostatistical tree — the hypothetical ancestor of the Nubian, Tama, and Nara languages. Among supporters of the East Sudanic and the broader Nilo-Saharan hypothesis, close relationship of these groups seems to be a given: it is supported by Lionel Bender (2005: 1), who groups these three taxa together into the "Ek" subbranch of East Sudanic (with the further addition of Nyimang), Christopher Ehret (2001: 88–89), who calls this tripartite taxon "Astaboran", and Claude Rilly (2009: 44), who agrees with Bender's classification, renaming his "Ek" subbranch "Northeast Sudanic" (as opposed to "Southeast Sudanic", comprising Surmic, Nilotic, and several other branches). However, a formal demonstration of this relationship based on a general, universally applicable methodology is still lacking, to the extent that some "conservative" encyclopaedic sources do not acknowledge the genetic link between these language groups as established beyond reasonable doubt<sup>9</sup>.

---

<sup>8</sup> Occasional biconsonantal matches can be found as well, but these are almost always scattered and confined to pairwise rather than multilateral matches — cf., for instance, a curious match between Temein and Ik in the word for 'star': Ik *dółéát* = Temein *qúli-t*, pl. *kó=qúl-á?* id. Considering that lexical contacts between speakers of Temein, who dwell in the Nuba mountains, and Ugandan Ik people are hardly likely, this phonetic similarity is currently best explained as an accidental resemblance.

<sup>9</sup> Cf.: "No conclusive, methodologically sound basis for assigning Nubian to East Sudanic or to an alleged full or partial Nilo-Saharan has been presented" (Hammarström et al. 2017: <http://glottolog.org/resource/languoid/id/nubi1251>).

The perfect way to demonstrate this relationship would have been a thorough, methodologically rigorous reconstruction of the phonological inventory of Proto-Nubian-Nara-Tama, supported by a large etymological corpus and based on recurrent phonetic correspondences, along with comparative grammatical evidence. However, even such a demonstration, in order to be easily appreciated by non-specialists in these languages, would still have to distinguish between “core” and “peripheral” layers of evidence, where only the “core” would serve the primary purpose of proving the relationship, whereas the “peripheral” layer (e.g. comparanda drawn from the cultural lexicon, featuring phonetic irregularities or questionable semantic shifts, etc.) would rather serve the purpose of multiplying our alleged knowledge on the already proven common ancestor of Nubian, Nara, and Tama.

Therefore, our intention here is to concentrate on the “core” evidence, extracting it by means of a formal lexicostatistical procedure. The procedure involves:

- demonstrating that a statistically significant number of phonetic homologies is detected between the compared protoforms for Proto-Nubian, Proto-Tama, and Nara equivalents for Swadesh meanings on the 50-item wordlist;
- interpreting these homologies in terms of regular phonetic correspondences, bringing in additional lexical data where necessary or possible;
- detecting additional potential cognates on the same wordlist that have not been identified automatically due to general limitations of the “consonantal class” method, and also interpreting them in terms of regular correspondences, if possible;
- detecting even more additional potential cognates between the compared taxa that involve typologically frequent, “trivial” semantic shifts from a basic Swadesh meaning to a semantically adjacent meaning;
- justifying a *genetic* rather than *areal* interpretation of the attested homologies/regulations by analyzing their distribution across various subdivisions of the 50-item wordlist, from terms that are “more stable on the average” to those that are “less stable on the average”.

### The data

Complete 100-item Swadesh wordlists have been compiled and annotated for all the languages from the three taxa in question where officially published or archival data were available in sufficient quantity; semantic selection of the optimal equivalents was performed based on the guidelines laid down in Kassian et al. 2010. Reconstruction of the optimal wordlists for Proto-Nubian and Proto-Tama (Nara, having no close relatives of its own, does not require a separate reconstruction, although one might occasionally resort to elements of internal reconstruction) was carried out for the 50-item subdivision of the complete 100-item wordlist; since a very detailed explanation for each of the items has already been published in Starostin 2013, only the least trivial and most significant decisions will be outlined in this paper.

Below we list all the principal data sources and briefly comment on the internal taxonomy of the respective language groups, as well as on previous and current research on the phonological reconstruction of their ancestral states.

A. *Nubian*. Wordlists were compiled for 10 languages belonging to the Nubian group:

(a) Nobiin; primary source — Werner 1987, with Bell 1970 used as an additional control source and Lepsius 1880 consulted for historical purposes. Unfortunately, the large dictionary Khalil 1996 may not be used for lexicostatistical purposes, since it intentionally omits all Arabic borrowings and mixes together data from a variety of old and new sources on different dialects of the language.

(b) Kenuzi-Dongolawi. These two closely related languages (or dialects of a single macro-language) are respectively represented by the data in Hofmann 1986 (Kenuzi) and Armbruster 1965 (Dongolawi), with Massenbach 1962 used as a control source for both.

(c) Hill Nubian. This large cluster of relatively small languages, scattered among the Nuba Hills, is represented by wordlists for Dilling (primary source: Kauczor 1920, with Jabr el Dar 2006 used for additional control), Kadaru, Debri (primary source: Thelwall 1978), Karko, and Wali (primary source: Krell 2012). Older data from Carl Meinhof's comparative vocabulary of Nubian languages (Meinhof 1918) have also been consulted for historical purposes, but are unusable as primary sources.

(d) Birgid; primary source — Thelwall 1977, with MacMichael 1920 consulted for control/historical purposes; since this language, constituting a significantly divergent branch of Nubian, has been reported as extinct, every bit of older data on it is extremely valuable.

(e) Midob; primary source — Werber 1993, with Thelwall 1983 consulted for control purposes.

In addition, a wordlist for the Old Nubian language, represented by texts from the 8<sup>th</sup> – 11<sup>th</sup> centuries A.D., has also been compiled based on the comprehensive dictionary of Gerald Browne (1996). Although the amount of recovered texts and their lexical content is large enough to permit the use of Old Nubian for lexicostatistical purposes, it has only been possible to fill in 75 out of 100 slots (and a few of these entries remain under serious doubt for various reasons), so any lexicostatistical conclusions on replacement rates between Old Nubian and modern Nubian dialects must be made with caution.

Worse still, although this topic has not been seriously explored so far, there are reasons to suggest that from a lexical perspective, “Old Nubian” is not a concise single dialect, but an amalgamation of several distinct speech varieties: thus, lexical analysis indicates every once in a while the presence of “doublets”, in which one word is cognate with its equivalent in modern Kenuzi-Dongolawi and the other one with the equivalent in modern Nobiin (e. g. *ŋul-* vs. *ado-* ‘white’, or *aman-* vs. *asse-* ~ *essi-* ‘white’). This goes against the general idea of Old Nubian as being specifically the ancestor of modern “Fadidja/Mahas”, i. e. Nobiin dialects (Browne 2002: 1), although from a formal statistical perspective, Old Nubian does have more in common with Nobiin than with Kenuzi/Dongolawi, and it makes more sense to assume a number of Kenuzi-Dongolawi interpolations in the Old Nubian corpus rather than to assign Old Nubian to a third separate subbranch of the Nile-Nubian branch (see below for more details on the overall classification of Nubian); this conclusion also agrees with the additional data on the varied nature of Old Nubian texts as adduced in Bechhaus-Gerst 2011: 20–22.

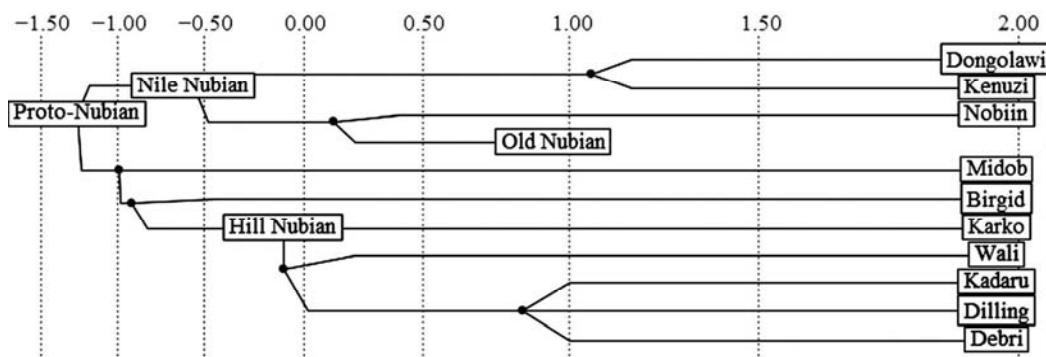
The main principle employed in the construction of a unified wordlist for Old Nubian has been that of statistic frequency. *Hapax legomena* or contextually ambiguous forms were accepted as main entries only in those cases where no other equivalents for the required Swadesh meaning were available. In case of “doublets” where one word is frequently encountered in texts and the other one is basically a *hapax*, only the frequently used word was included in the calculations. Consistent use of this principle showed that the majority of exclusive isoglosses, as a result, is indeed between Old Nubian and Nobiin rather than Old Nubian and Kenuzi-Dongolawi.

Refined lexicostatistical calculations (slightly revised and corrected as compared to the previous analysis in Starostin 2014: 34) yield the following percentage matrix for Nubian (Table 2), which, through the application of Sergei Starostin's revised glottochronological method and the Starling-NJ phylogenetic method (Burlak, Starostin 2005: 162–167; Kassian 2015), may then be converted to the following tree format (Figure 2).

Table 2. Lexicostatistical matrix for Nubian languages (100-item Swadesh wordlists)

|     | NOB  | DNG  | KNZ  | DIL  | KAD  | DEB  | KRK  | WLI  | BIR  | MID  |
|-----|------|------|------|------|------|------|------|------|------|------|
| ONU | 0.81 | 0.63 | 0.63 | 0.42 | 0.43 | 0.44 | 0.45 | 0.42 | 0.39 | 0.51 |
| NOB |      | 0.66 | 0.66 | 0.40 | 0.42 | 0.41 | 0.41 | 0.39 | 0.42 | 0.51 |
| DNG |      |      | 0.93 | 0.59 | 0.61 | 0.62 | 0.55 | 0.54 | 0.56 | 0.57 |
| KNZ |      |      |      | 0.60 | 0.59 | 0.60 | 0.55 | 0.55 | 0.56 | 0.57 |
| DIL |      |      |      |      | 0.92 | 0.91 | 0.75 | 0.76 | 0.64 | 0.57 |
| KAD |      |      |      |      |      | 0.92 | 0.79 | 0.81 | 0.60 | 0.56 |
| DEB |      |      |      |      |      |      | 0.80 | 0.82 | 0.59 | 0.57 |
| KRK |      |      |      |      |      |      |      | 0.72 | 0.56 | 0.53 |
| WLI |      |      |      |      |      |      |      |      | 0.59 | 0.55 |
| BIR |      |      |      |      |      |      |      |      |      | 0.56 |

Figure 2. Phylogenetic tree for Nubian languages (with glottochronological interpretation)



This classification largely agrees with the traditional model as described, e.g., in Bechhaus-Gerst 1985, with a rapid disintegration of Common Nubian into four different branches (Nile-Nubian, Midob, Birgid, and Hill Nubian), but sharply contradicts the later reclassification in Bechhaus-Gerst 1989 and 1996; according to Bechhaus-Gerst, Nobiin should be excluded from Nile-Nubian and positioned as the first branch to split off from Common Nubian, while the increase in lexical and grammatical similarity with Kenuzi-Dongolawi is explained by her as the result of a prolonged period of convergence. This re-classification has been critically scrutinized in Starostin 2014: 93–96, and still more recently in Vasilyev, Starostin 2014, where it was concluded that Nobiin is indeed far more lexically divergent from the rest of Nubian than any other constituent of this group, but that the divergent elements are consistently better interpreted as representing a non-Nubian substrate rather than archaisms inherited from Proto-Nubian<sup>10</sup>; subsequently, the convergence phenomenon must have taken place between Nobiin and some non-Nubian language or languages that used to be spoken to the north of the original Nubian homeland, rather than between Nobiin and Kenuzi-Dongolawi. Results of the analysis convince us that there is no need to dismantle the old Nile-

<sup>10</sup> Precisely the same conclusion has been independently reached by Claude Rilly (2009: 285–288).

Nubian branch, but that there is every reason to treat Nobiin data with caution when it comes to external comparison, particularly if it finds no parallels in other Nubian languages.

The first attempt to establish regular phonetic correspondences between various Nubian languages and set up a Proto-Nubian reconstruction was carried out by Ernst Zyhlarz (1950), but the research was largely inadequate due to lack of sufficient data sources on Hill Nubian, Birgid, and Midob. The first truly significant reconstruction of the Proto-Nubian phonological system, supported by a small etymological vocabulary and still fully relevant today, was carried out by Marianne Bechhaus-Gerst (1985); since then, a somewhat more refined version has been offered by Claude Rilly (2009: 211–288), and additional observations on the complex developments of Proto-Nubian phonology in Hill Nubian languages were made by Angelika Jakobi (2006). The reconstruction system adopted in Starostin 2014 and, consequently, this paper as well, rests largely on the research of Bechhaus-Gerst, but offers a few corrections, for the most part, concerning non-standard consonantal behavior in clusters that appear on morphemic borders; some of these are briefly commented upon below in connection with specific items. In most of the proposed systems, Nile-Nubian languages (and possibly also Birgid) are generally viewed as more phonologically conservative, but data from Hill Nubian and Midob are also essential in order to better assess the distribution of cognates in daughter branches and make more reliable choices for Swadesh meanings on the Proto-Nubian level.

B. *Nara* (= Barea). Nara is typically described as a linguistic isolate, although sources note that the language may be divided in at least two distinct pairs of dialects: Eastern (Higir-Mogoreeb) and Western (Koyta-Saantoorta), with limited mutual intelligibility (Rilly 2005: 1, 2009: 178). Unfortunately, all available sources of significant data concentrate exclusively on Higir as the most widely spoken variety of Nara, which leaves no space for a serious historical reconstruction. The most important of these are Bender 1968, with a 200-item wordlist, and the much earlier descriptive monograph by Leo Reinisch (1874), which also contains a detailed vocabulary. For etymological research, the somewhat later grammatical sketch Thompson 1976 and a few recent works, like Hayward 2000 on the Nara tonal system or Abushush, Hayward 2002 on general phonology, also provide some limited data support.

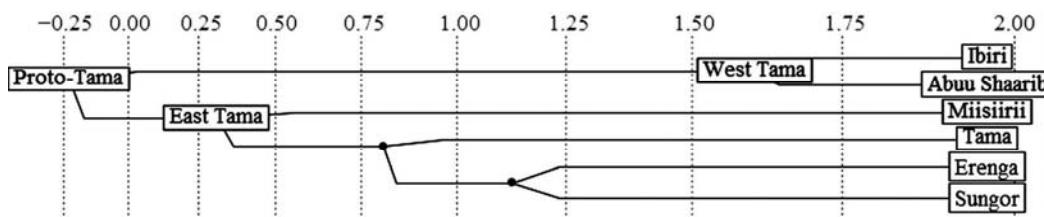
C. *Tama*. Descriptive work on this small, but significantly diversified language group, spoken in Ouaddaï and Dar Fur, has been very scarce so far, with no grammars or dictionaries produced for even a single language. The principal source of data, in fact, remains officially unpublished: it is a comparative vocabulary of all known Tama languages, compiled by John Edgar (1990) from the largest possible variety of sources, including his own field data as well as records stretching all the way back to the late 19th century, and also incorporating data from printed sources such as Lukas 1933 on Ibiri and Lukas 1938 on Sungor. Although made available (by kind courtesy of Roger Blench) in almost print-ready form, the work formally retains the status of a manuscript due to the author's untimely demise; only a few bits of the data appeared in print form, illustrating Edgar's pioneering attempt at a reconstruction of Proto-Tama phonology (Edgar 1991a).

According to Edgar's classification that has also been lexicostatistically confirmed in Starostin 2014, the Tama group is divided into two primary branches: the smaller West Tama cluster, consisting of Ibiri (Mararit) and its satellite dialects such as Abu Sharib, and the larger East Tama cluster, which is itself divided into Miisiirii and Tama-Erenga-Sungor (three closely related dialects). Data for all five varieties, collected in Edgar 1990, are sufficient to construct near-complete Swadesh wordlists that yield the following cognacy matrix (Table 3; also slightly revised as compared to the previous analysis in Starostin 2014: 317), and the following phylogenetic tree (Fig. 3; also constructed by means of the Starling-NJ method).

Table 3. Lexicostatistical matrix for Tama languages (100-item Swadesh wordlists)

|     | ERE  | SUN  | MIS  | IBI  | ASH  |
|-----|------|------|------|------|------|
| TAM | 0.89 | 0.91 | 0.80 | 0.69 | 0.71 |
| ERE |      | 0.94 | 0.85 | 0.69 | 0.69 |
| SUN |      |      | 0.85 | 0.70 | 0.68 |
| MIS |      |      |      | 0.70 | 0.67 |
| IBI |      |      |      |      | 0.99 |

Figure 3. Phylogenetic tree for Tama languages (with glottochronological interpretation)



It is important to note that Tama gives the (glottochronologically confirmed) impression of a less chronologically deep family than Nubian; consecutively, its 50-item proto-wordlist is easier to reconstruct due to fewer lexical replacements in the principal branches. Nevertheless, some of the languages have still gone through significant phonetic change, not all of which is easy to trace and reliably reconstruct due to limited (and not always accurately transcribed) amounts of data. Our reconstruction of Proto-Tama depends significantly on the rules laid down in Edgar 1991a, with some additions and corrections offered in Starostin 2014: 314–316.

### Comparative 50-item wordlists for Proto-Nubian, Nara, and Proto-Tama.

*Preliminary notes.* Table 4 below does not list the complete data (freely available at the website of the Global Lexicostatistical Database), but only the reconstructed optimal candidates for 50 out of 100 semantically fixed “Swadesh slots” (detailed explanation of semantics for each slot may be found in Kassian et al. 2010) for Proto-Nubian and Proto-Tama; Nara is represented by Higir dialect data from Bender 1968. Numeric indexes that follow individual items reflect their average “stability index” as per Starostin 2010: 113 (ultimately based on the calculations across various genetic lineages in Eurasia, Africa, and Australia as per S. Starostin 2007).

Detailed justifications for all the reconstructions may be found in Starostin 2014; in this paper, due to volume considerations, notes on particular reconstructions will be condensed and restricted to non-trivial cases of phonetic or semantic developments, while the majority of the notes section will concentrate on the justification of etymological matches between PN, PT, and Nara.

We use the following notation symbols to designate various degrees of cognacy estimation:

! — marks pairs or triplets of reconstructions whose simplified phonetic shapes (“consonantal skeletons”) match each other according to the Dolgopolsky consonantal class criterion. In cases where two or more reconstructions are more or less equiprobable for one taxon (either because there is no certainty about the phonetic interpretation of a given proto-etymon, or because two different etyma are represented in two primary branches of the family), in the table below we only list the variant that is compatible with potential external cognates.

+ — marks pairs or triplets of reconstructions that represent highly probable etymological cognates. Although at this point, despite the works of M. L. Bender and C. Rilly, it is probably too early to talk about a definitive set of regular phonetic correspondences for East Sudanic as a whole or Northeast Sudanic (Nubian-Nara-Tama) in particular, we provisionally mark the forms as cognate with each other if the consonantal correspondences between them are trivial (i.e. the consonants are exactly the same) or may be shown to form a part of a recurrent pattern (e.g. Proto-Nubian *\*n-* = Proto-Tama *\*l-*) or may be explained as the result of morphophonological or morphological processes. Precise vocalic correspondences are not expected, but the base root vowels should have a certain degree of proximity, i.e. a match between labial vowels *\*o* and *\*u* is acceptable, while a match between *\*a* and *\*i* is suspicious. Predictably, there will be a serious correlation rate between “automated” and “etymological” cognates, but not a 100% one (see ‘drink’, ‘egg’, etc.).

[] — square brackets mark items that have neither “automated” nor “etymological” parallels in any of the other two groups.

’ — this special symbol is typically inserted after the initial vowels of VCVC-type stems, typically encountered in Proto-Nubian, more rarely in Nara, and almost never in Proto-Tama. Since the most common type of root structure for all these languages is CVC, this initial vowel, often identical in quality to the main root vowel (cf. in Proto-Nubian: *\*ubur-* ‘ashes’, *\*awar-* ‘night’, etc.; there are, however, exceptions such as *\*agul-* ‘mouth’, etc.), may be suspected of representing an old fossilized prefix, perhaps the trace of one or more older classifiers or determinants, which justifies its formal deletion in the procedure of external comparison. Alternatively, this vowel may have been an integral part of the original root, in which case it would be possible to regard the Proto-Nubian system as more archaic in comparison with Nara and Tama, where it became lost due to purely phonetic processes.

Table 4. 50-item wordlist entries for Proto-Nubian, Nara, and Proto-Tama.

| #  | Word                      | Proto-Nubian       | Nara     | Proto-Tama           |
|----|---------------------------|--------------------|----------|----------------------|
| 1  | ‘ashes’ <sup>38</sup>     | <i>*u'bur-ti</i> + | hibid ?  | <i>*or-ŋo</i> +      |
| 2  | ‘bird’ <sup>33</sup>      | <i>*kawir-</i> +!  | karba +! | [*wig-]              |
| 3  | ‘black’ <sup>48</sup>     | [ <i>*u'dum-</i> ] | [sur-ku] | [*kidi-]             |
| 4  | ‘blood’ <sup>20</sup>     | [ <i>*a'ger</i> ]  | [kito]   | [*ya-i]              |
| 5  | ‘bone’ <sup>34</sup>      | <i>*kəsi-di</i> +  | ketti +  | <i>*ki-(ŋa)-ti</i> + |
| 6  | ‘claw/nail’ <sup>19</sup> | <i>*sunj-di</i> ?  | ši       | <i>*ŋosa-</i> ?      |
| 7  | ‘die’ <sup>13</sup>       | <i>*di:-</i> +!    | di:- +!  | [*iye] (← Maba?)     |
| 8  | ‘dog’ <sup>16</sup>       | <i>*bəl</i> ?      | wəs +!   | <i>*wes-i</i> +!     |
| 9  | ‘drink’ <sup>15</sup>     | <i>*ni:-</i> +     | li:- +!  | <i>*li-</i> +!       |
| 10 | ‘dry’ <sup>24</sup>       | [ <i>*sow-</i> ]   | [dise-]  | [*lab-]              |
| 11 | ‘ear’ <sup>32</sup>       | <i>*ulgi</i> ?     | tus ?    | <i>*ŋ=us</i> ?       |
| 12 | ‘eat’ <sup>25</sup>       | <i>*kɔl-</i> +!    | kʌl- +!  | [*ŋan-]              |
| 13 | ‘egg’ <sup>47</sup>       | <i>*kumbu</i> +    | [wari]   | <i>*kob-</i> +       |
| 14 | ‘eye’ <sup>4</sup>        | <i>*mijŋ</i> +!    | [no]     | <i>*e'mep-</i> +!    |
| 15 | ‘fire’ <sup>7</sup>       | <i>*usi-gi</i> +!  | ši-ta ?  | <i>*us-g</i> +!      |
| 16 | ‘foot’ <sup>43</sup>      | [ <i>*oy</i> ]     | [bəla]   | [*war]               |
| 17 | ‘hair’ <sup>27</sup>      | [ <i>*dɛl-</i> ]   | [sebi]   | [*isigi-t]           |
| 18 | ‘hand’ <sup>11</sup>      | <i>*θ-si</i> +     | a:(-)t + | <i>*aw-g</i> +       |

| #   | Word                  | Proto-Nubian    | Nara          | Proto-Tama    |
|-----|-----------------------|-----------------|---------------|---------------|
| 19  | ‘head’ <sup>49</sup>  | *or +!          | [kela]        | *ur +!        |
| 20  | ‘hear’ <sup>45</sup>  | *giʒ- ?         | [wos /Rn./]   | *sig- ?       |
| 21  | ‘heart’ <sup>14</sup> | [*ay-]          | a’sim-a +!    | *samil +!     |
| 22  | ‘horn’ <sup>44</sup>  | *ŋəʒi +         | [keli]        | *ŋawi-ti +    |
| 23  | ‘T <sub>3</sub>       | *ə-y +!         | a-g +!        | [*wa]         |
| 24  | ‘kill’ <sup>42</sup>  | [*pay-]         | si- +!        | *siy- +       |
| 25  | ‘leaf’ <sup>41</sup>  | [*ulgi]         | [tifini]      | [*afol]       |
| 26  | ‘louse’ <sup>17</sup> | [*i/ŋ/-ti]      | ši-ti +       | *sin- +       |
| 27  | ‘meat’ <sup>46</sup>  | [*kosi]         | [nɔ:-]        | [*is-]        |
| 28  | ‘moon’ <sup>18</sup>  | [*ŋun-]         | [fe:ta]       | [*ayi-]       |
| 29  | ‘mouth’ <sup>31</sup> | *a’gul +        | a’wɔlo +      | *kul +        |
| 30  | ‘name’ <sup>10</sup>  | [*ə’ri]         | a:da +!       | *at +!        |
| 31  | ‘new’ <sup>23</sup>   | [*ɛ:r]          | [wər- ~ wɔr-] | [*suw-]       |
| 32  | ‘night’ <sup>50</sup> | *a’war +!       | [kriš-]       | *war +!       |
| 33  | ‘nose’ <sup>29</sup>  | [*esi-ŋ(i)]     | [dəmмо]       | [*miʒi]       |
| 34a | ‘not’ <sup>30</sup>   | *m- +!          | ma= +!        | *m= +!        |
| 34b | ‘not’ <sup>30</sup>   | [*=a-]          | [ka=]         | [*=to]        |
| 35  | ‘one’ <sup>21</sup>   | [*bey-]         | [doku]        | [*ku- ~ *ka-] |
| 36  | ‘rain’ <sup>39</sup>  | *ar- +!         | [hala]        | *ar- +!       |
| 37  | ‘smoke’ <sup>36</sup> | [*gume-]        | [a’suru]      | [*turu-]      |
| 38  | ‘star’ <sup>40</sup>  | *wane +!        | wi:ni +!      | *mij- +       |
| 39  | ‘stone’ <sup>9</sup>  | [*kul-]         | [ta:na]       | [*kad-]       |
| 40  | ‘sun’ <sup>35</sup>   | [*maša-]        | [ko:s]        | [*ari]        |
| 41  | ‘tail’ <sup>26</sup>  | [*ɛ:b]          | [dawa]        | [*gawu-t]     |
| 42  | ‘thou’ <sup>5</sup>   | *e- ~ *i- +!    | i-ŋa +!       | *i- +!        |
| 43  | ‘tongue’ <sup>8</sup> | *ŋalT- +        | [haga]        | *laja-t +     |
| 44  | ‘tooth’ <sup>22</sup> | [*ŋəl-]         | nihi +        | *ŋeʒ- +       |
| 45  | ‘tree’ <sup>37</sup>  | [*pər]          | [*kel]        | [*ga:n]       |
| 46  | ‘two’ <sup>2</sup>    | *awri +!        | ari +!        | *wari +!      |
| 47  | ‘water’ <sup>28</sup> | [*əs-]          | [mba]         | [*ka:l]       |
| 48  | ‘we’ <sup>1</sup>     | *a-y +!         | a-gga +!      | [*wa-i]       |
| 49  | ‘what’ <sup>12</sup>  | *nwa- ~ *nwi- + | [nda-]        | *num +        |
| 50  | ‘who’ <sup>6</sup>    | [*ŋə-y]         | na- +!        | *na +!        |

*Comments on individual entries.*

1. ‘Ashes’. PN \*u’bur-ti (Nob. ūbúr-tí, Dng. ub’ur-ti, Knz. ubur-ti; Dil. ɔp-te, Kad., Deb. ɔt-te, Krk. ōmì-t; Bir. ubur-ti; Mid. úfù-di) = PT \*or-ŋo (Ere. orŋo, Sun. orŋo ~ oruŋo, Mis. arŋo).

The element \*-ŋ- in the PT form is easily analyzable as a fossilized plural/collective suffix (the same morpheme is frequently found as a productive pluralizer as well). Root morpheme \*or- is derivable through lenition and contraction from an earlier \*owur- ← \*obur-; for similar cases of possible development of labial \*b before labial vowels cf., e.g., PN \*u’bur ‘hole’ = PT \*war- ~ \*wor- id. (although here PT probably reflects a variant without the prefixal vowel).

Nara *hibid* (Bd.), *hübet* (R.) could also belong here, provided the *h*- is prosthetic and the word-medial cluster has been simplified (\**ubur-ti* → \**hubir-ti* → \**hubit*); however, this is a complicated scenario that needs additional evidence, so we cannot count this as a *bona fide* match.

2. ‘**Bird**’. PN \**kawar-ti* (ON *kawar-t-*; Nob. *kawar-ti*, Dng. *k'awir-te* ~ *k'auir-te* ~ *k'aur-te*, Knz. *kawir-te*; Dil. *kəmil-ti*, Krk. *kübür-àn*; Bir. *kwar-ti*; Mid. *à:béd-dí*) = Nara *kərba* (Bd.), *karba* (R.).

Phonetically compatible under the assumption of a metathesis in Nara (\**kawər-* → \**karb-*), which seems typologically plausible and finds no contradictory evidence.

PT \**wig-* ‘bird’ (Tama *wigí-t*, Ibi. *wígí-t*, etc.) is incompatible with these forms and finds no obvious parallels in either PN or Nara.

3. ‘**Black**’. No parallels detected between any of the three taxa.

4. ‘**Blood**’. No parallels detected between any of the three taxa.

5. ‘**Bone**’. PN \**kəsi-di* (ON *gis-ri-*; Nob. *gisi-r*, Dng. *kh'i:-d*, Knz. *ki:-d*; Kad. *kwe-de*, Deb. *kwe-du*, Krk. *kwīè-dò*, Wal. *kwī-tù*; Bir. *kizí-di*; Mid. *à:-dí*) = Nara *kə-ti* (Bd.), *ke-tti* (R.) = PT \**ki-(ŋa)-ti* (Tama *kí-tí*, Ere. *kīŋà:-tí*, Mis. *kiŋi-t*, Ibi. *kiŋí-t*).

In Starostin 2014: 320 it was suggested that the PT paradigm should be reconstructed as \**kiŋa-ti* (sg.), \**kiŋa-k* (pl.), with vowel reduction and cluster simplification in Tama proper: \**kiŋa-ti* → \**kiŋ-ti* → *ki-ti*. However, since then I have found no corroborative evidence for the latter development; and considering the relative frequency of *-ŋ-* as a plural marker in Tama languages, it is perfectly plausible to reinterpret this as sg. \**ki-ti*, pl. \**ki-ŋa*, with subsequent generalization of the plural form in most Tama languages and reformation of the entire paradigm based on it (with new singulative \**ki-ŋa-ti* and new plural \**ki-ŋa-k*).

This interpretation is in good agreement with Nara data, suggesting a common Tama-Nara root \**ki-* or \**kə-*. The parallel with Nubian is slightly more problematic, but intervocalic \**-s-* on the whole is a fairly unstable consonant in this entire region (cf. lenition and elision in Kenuzi-Dongolawi for this very root, or the regular deletion of \**-s-* in East Tama languages), so the assumption of a regular development \**kəsi-ti* → \**kə-ti* ~ \**ki-ti* in Tama and Nara, even without additional evidence for the moment, seems fairly realistic. In any case, at least the Nara-Tama isogloss is unquestionable.

6. ‘**Claw/nail**’. PN \**sunj-di* (Nob. *sun-ti*, Dng. *sun-ti*, Bir. *sunj-di*, etc.) and PT \**ŋosa-* (Mis. *ŋɔsə-t*, Sun. *ŋisi-t*, etc.; see Starostin 2013: 320–322 for a detailed discussion on the complicated fate of this etymon due to its contamination with ‘tooth’ in the individual languages) may actually be relatable to each other through metathesis, although it is impossible to say which form should be thought of as representing the original consonantal sequence. However, since this kind of metathesis would have to be qualified as an incidental irregularity, it is difficult to count this parallel as a primary piece of etymological or lexicostatistical evidence for the Nubian-Tama relationship.

7. ‘**Die**’. PN \**di:-* (ON *di:-*, Knz.-Dng. *di:-*, Mid. *tí:-*, etc.) is a perfect match with Nara *di:-*. No sign of this root appears in Tama, and, in fact, Tama \**iye* is one of the few entries on the list which, instead, shows close phonetic proximity to Maba languages, cf. Masalit *iy*, Kibet *iy*, Kodoi *yî*, Maba *óy* ‘to die’ (Edgar 1991b: 391). Borrowing from Maba is not the only possibility (similar forms are also found on some proto-levels in other East Sudanic languages, e. g. East Nilotc \**=ye-* ‘to die’), but, in any case, it is impossible to relate the Tama equivalent to Nara and/or Nubian.

8. ‘**Dog**’. Nara *wəs* (Rn.: *wos*) is clearly the same as Tama \**wes-i* (Ib. *wí:si*, AS *wis*, Mis. *wus*; Tama *wéi*, Ere. *wi*, Sun. *we*: with regular deletion of intervocalic \**-s-*). On the possibility of Nubian \**bəl* (Dng. *wəl*, Dil. *bol*, Bir. *mel*, Mid. *pə:l*, etc.) being related to Tama and Nara through a non-trivial consonantal correspondence see below (‘ear’); at present, however, we prefer to keep these etyma apart.

9. ‘**Drink**’. The obvious parallel is between Nara *li*- and Proto-Tama \**li/y*- (Tama *li*; Ere. *li*-ε, Sun. *liy*-ε, Mis. *liy*-*ei*, AS *li*, etc.). However, both forms also regularly correspond to Proto-Nubian \**ni*- (ON *ŋi*-, Dng. *ni*; Dil. *di*, Bir. *ni*; Mid. *ti*-, etc.): Proto-Nubian has no word-initial \**l*-, which makes the assumption of regular development \**n*- → \**l*- perfectly plausible, and furthermore, the correspondence may be strengthened by additional examples, even from the basic lexicon (e. g. Proto-Tama \**lasi*- ‘long’ = Proto-Nubian \**nas*- ‘long’).

10. ‘**Dry**’. No parallels. This is not a stable item in either Nubian or Tama (most sub-branches have their own replacements, and precise reconstruction is very difficult).

11. ‘**Ear**’. This is a complicated case where additional progress might be made in subsequent etymological studies of the Nubian-Nara-Tama family.

In Tama, the root is \**us*- (Tama sg. *ú-tù*, pl. *ú-ŋ-ðŋ*, Ere. sg. *us-ut*, pl. *us-ŋŋ*, Mis. sg. *us-ut*, pl. *us-ŋŋ*), but in West Tama, it is preceded by a fossilized prefixal element *ŋ*=: Ibi. *ŋùs-i*, AS *ŋgus-i*. The function of this prefix remains obscure, yet its segmentable status is corroborated quite firmly by additional examples (see ‘head’ and ‘name’ below). The ability of the root \**us*- to combine with a fossilized prefix reasonably begs the question of whether a *different* fossilized prefix (with an equally obscure function) could not be present in Nara *t(=)us* ‘ear’. However, unlike Tama, in Nara no additional evidence has been found so far to suggest the idea of a formerly segmentable *t*=; in fact, the only other basic lexicon term with initial *t*- that shows credible outside parallels is Nara *tawa* ‘belly’ = Proto-Nubian \**tu* id., without any signs of segmentation. Therefore, this comparison remains highly questionable and unfit as primary evidence for relationship.

A different problem is tied to Proto-Nubian \**ulgi* ‘ear’ (ON *ulgi*-, Dng. *ulug*, Nob. *úkkí*, Mid. *úlgí*, etc.). If we assume that the second syllable is of suffixal origin, the allegedly original root \**ul*- would correlate with Tama \**us*- precisely the same way that Proto-Nubian \**bal* ‘dog’ (see above) correlates with Nara *wəs*, Tama \**wes-i* — suggesting a non-trivial correspondence «Proto-Nubian \**l* : Proto-Tama \**s* : Nara *s*» whose most logical phonetic interpretation would be a lateral fricative (\**ɬ*). This idea seems worthy of further exploration, but for the moment, no further examples of this correspondence are available, and we cannot qualify either of these parallels as primary evidence.

12. ‘**Eat**’. PN \**kal*- (Dng. *kal*, Dil. *kol*, Mid. *ɬl*-, etc.) is perfectly compatible with Nara *kal*- (Rn.: *kal*-). The Tama paradigm is completely different: East Tama \**ŋan*- is opposed to West Tama suppletive forms: imperfective \**gey*- vs. perfective \**sin*- . None of the three forms has anything to do with the verb in Nara or Tama.

13. ‘**Egg**’. Nile-Nubian \**kumbu* (ON *kumpu*-, Dng. *kumbu*, Nob. *kúmbú*), one of several equiprobable candidates for PN ‘egg’, is comparable with West Tama \**kob*- (Ibiri *kób-it*, AS *ka:b-it*), assuming cluster simplification in the latter (nasal cluster *-mb*- does not seem to be encountered in inherited lexicon in these languages).

14. ‘**Eye**’. Some of the phonetic shapes in Nubian and Tama languages are almost completely identical, cf. Dng. *missi* vs. AS *meše* (Barth), etc. However, detailed etymological analysis of the complete datasets, as presented in Starostin 2014: 50–51 (for Nubian) and 328–329 (for Tama), shows that in both of these groups, there is serious evidence for reconstructing a “weak” palatal nasal in root-final position, prone to elision or assimilation — but still preserved in some Nubian languages (ON *map*-, Nob. *má:p*) and, in assimilated form, in such relic plural forms as Ibiri *imn-ién* ← \**e=meŋ-oŋ*. This means basic compatibility for the reconstructed variants as well, allowing us to posit \**mij*- ~ \**meŋ*- as the optimal equivalent for ‘eye’ on the proto-level.

In comparison, Nara *no* ‘eye’ shows no affinity with these forms, but it makes sense to compare the Nubian and Tama items with the Nara verb *minni* ‘to flash, shine’ (Rn.): if the

etymology is correct, we could be dealing with a Nubian-Tama shared innovation ('to shine' → 'eye') vs. a possibly retained archaism in Nara.

15. 'Fire'. Here we have a transparent isogloss between Nubian *\*usi-gi* (cf. especially Bir. *uzug* and Mid. *ússí*; such forms as Old Nubian *eig-*, Nobiin *i:g*, etc., probably represent contractions of the original stem) and Tama *\*us-g* (Ibiri *ùsùg-i*, AS *usugu*; Tama *ú*, Ere. *ú*, etc., are also contracted variants, with regular deletion of intervocalic *\*-s-* in these languages). It is unclear if Nara *šita* 'fire' also belongs here, but it is possible: *-ta* may be identified as a fossilized plural suffix (cf. Nara *nó:-ta* 'meat' [Bd.] vs. the earlier recorded *no* [Rn.]), and the word-initial vowel could be syncopated in a trisyllabic structure (unless it was a detachable prefix from the very beginning). However, both of these assumptions remain rather speculative.

16. 'Foot'. No parallels detected between any of the three taxa.

17. 'Hair'. There is distinct phonetic similarity between Tama *\*isigi-* (Tama *igí-t*, Ere. *sigi*, Ibiri *isíñjí-t*, AS *isíñjí-t*) and Nobiin *šigír-tí* 'hair'. However, the latter, even if it is related (with irregular deletion of the word-medial consonant) to Knz. *sí:r* 'hair', is far from the optimal carrier of the basic meaning 'hair' in Proto-Nubian. Additionally, its phonetic proximity to various Semitic and Cushitic terms for 'hair' (e. g. Arabic *šafr-*, Ethiosemitic *\*sag<sup>w</sup>ar*, etc.) makes all these items highly questionable as potential genetic markers, so we would not want to consider them as primary evidence.

18. 'Hand'. All compared forms may be regarded as cognates, although phonetic similarity between them is obscured by the tendency of the original short root to get fused with various suffixes, formerly (or, sometimes, still productively) denoting singulative or plural semantics.

For PN, Rilly (2009: 477) reconstructs *\*es-i* 'hand', which almost coincides with *\*əsi* in Starostin 2014: 54; this form is either preserved with minimal phonetic change (Dil. *iši*, Mid. *àssi*), or is subject to regular weakening and deletion of intervocalic *\*-s-* (Dng. *i:*), or becomes further extended with an additional singulative marker (Nob. *èddì* ← *\*əsi-ti*). For PT, the suggested reconstruction is *\*awg* (Starostin 2014: 332), which seems to explain the wide variety of reflexes (Tama *àù*, Ere. *auw* ~ *oy*, Sun. *ao*, Mis. *wi*; Ibi. *wèi*) somewhat better than Rilly's *\*(a)wei* (2009: 477), although ultimately the basic consonantal shape of the reconstruction is the same in both cases, since we regard *\*aw-g* as a transitional fusion of the original root *\*aw(i)-* with a former plural marker.

All three forms, including Nara *a:t*, can be rather unproblematically traced back to an original root *\*ay-*, or, perhaps, a bisyllabic stem *\*ayi*, with the Proto-NNT paradigm *\*ayi-ti* (sg.) : *\*ayi-k-* (pl.) conforming to the very common so-called "T/K pattern" of East Sudanic (Bryan 1959). As both forms underwent contraction and fusion in daughter branches, only the first one survived in Nara (*\*ayi-ti* → *a:t*) and in Nubian, where assimilation with the fricative *\*-y-* resulted in fricativization of the old stop (*\*ayi-ti* → *\*ayti* → *\*ə(s)si*); PT, on the other hand, generalized the plural form, and, in addition, underwent a dissimilative process: *\*ayi-k-* → *\*awi-k-* → *\*awg-*. This dissimilation is precisely the same as in the case of 'horn' (see below) and may be considered regular.

Although short monoconsonantal stems beset with idiosyncratic issues of morphological fusion could be regarded as questionable evidence for genetic relationship, in this particular case it is worth noting that the word 'hand' also displays very similar patterns of behaviour in other potential East Sudanic languages as well; cf., for instance, the situation in East Nilotic, where the old root *\*k=ay-* (extended by means of the common nominal prefix *k=*) is still occasionally encountered as a segmentable unit (e. g. sg. *n=káí-ná*, pl. *n=káí-k* in Camus), but generally tends to fuse, once and for all, with the old singular marker *-n* (e. g. sg. *á=kán*, pl. *ŋá=kán* in Turkana, etc.; see Vossen 1982: 326 for more data). Similar situations are attested in Surmic,

Daju, Nyimang, and Temein: all these groups share the common invariant  $*a(y)-C-$  ~  $*e-C-$  for the meaning ‘hand’, where  $-C-$  is sometimes fused with the old root and sometimes remains as a productive number marker. These external parallels should certainly raise the level of confidence in the correctness of this Nubian-Nara-Tama etymology.

19. ‘**Head**’. Here we have a clear correlation between Nubian  $*or$  (→ ON *ur-*, Nob. *ùr*, Dng. *ur*, Mid. *òr*, etc.) and Tama  $*ur$  (→ Ibiri *úr-i*, AS *ur*; other Tama languages show an initial  $\eta=$  which must be some sort of fossilized, possibly pronominal or deictic, prefix — Tama *ηúr*, Mis. *ηor*, etc.; the same prefix is also encountered in ‘name’, see below).

Nara *kela* certainly does not belong here, but has a phonetically perfect and semantically acceptable parallel in Nile Nubian  $*kel-$  ‘end, border, tip’ → ON *kel-*, Dng. *ké:l*, Knz. *ke:l*, suggesting a semantic shift in Nara (‘tip, end’ → ‘head’) with loss of the original root.

20. ‘**Hear**’. PN is reconstructed based on an isogloss between Kenuzi-Dongolawi  $*giz-$  and such Hill Nubian forms as Dilling *ki-er-* (←  $*giz-$  with regular devoicing of initial velar and possibly regular loss of intervocalic  $-\mathfrak{z}-$ , although this has not been properly confirmed yet). Phonetic similarity of this stem with PT  $*sig-$  (Tama *ik-*, Sun. *ig-*, Mis. *sug-o*) is observable, but the two could be related only under the assumption of a spontaneous metathesis (cf. a similar possible metathesis between a velar and an alveolar consonant, but with reverse direction, in the case of ‘nail’), therefore, we should not accept this evidence as primary.

The Nara equivalent is incompatible, but if initial *w-* is prothetic, the verb *wos-* may actually contain the same root as Tama  $*\eta=us$  ‘ear’ (and even *tus* ‘ear’ in Nara itself, see notes on 11 ‘ear’ above). If so, this would be the same type of development as in Old Nubian *ulg-ir-* ‘to hear’, Nobiin *úkké-ér* id., a verbalization of PN  $*ulgi$  ‘ear’.

21. ‘**Heart**’. We reconstruct the PT form as  $*samil$  based on Mis. *samil* and forms with regular deletion of  $*s-$  in East Tama (Tama *àmúl*, Ere. *àmòl*, Sun. *amul*). Since triconsonantal roots in East Sudanic languages are a rarity, it is plausible to assume that  $-(i)l$  here is a fossilized suffix, same as the one that also occurs in some other nominal stems (e. g. Tama *tó-l-òl* ‘belly’ = Sun. *to-l* id., further perhaps to PN *\*tu* ‘belly’ without this marker) and possibly of the same origin as the Common Nubian determinant  $*-l$ . This allows easy comparison with Nara *asima*, at least as far as the basic consonantal skeletons are concerned. Some Nubian forms also show a stem with a fossilized determinant (PN  $*ay-il-$  → ON *ai-l-*, Dil. *a-l-du*, etc.), but the root proper is  $*ay-$  (→ Nob. *áy*, Bir. *ai-di*, etc.), not comparable with Nara and PT.

22. ‘**Horn**’. PN  $*\eta\mathfrak{z}i$  → Nob. *nì:šì*, Dng. *nišši*, Dil. *dɔ̄-ti* (regular development  $*\eta-$  →  $*n-$  → *d*- and probably the same regular deletion of  $-\mathfrak{z}-$  as in ‘hear’ q.v., with a new productive marker added), Bir. *ηis-ti*, Mid. *kó:çí*. We may plausibly interpret the form  $*\eta\mathfrak{z}i$  as a contraction from an older  $*\eta\mathfrak{ay}-ti$ , the same way that  $*\mathfrak{asi}$  ‘hand’ is contracted from  $*\mathfrak{ay}-ti$  (the only difference being that this time around, the initial voiced consonant caused the word-medial consonant to become voiced as well).

PT  $*\etaawi-ti$  is reconstructed based on Tama *ηó-d* (pl. *ηó-n*), Ere. *ηε-ti*, Sun. *ηο:-tu*; with the same dissimilation as in ‘hand’ ( $*ayi-$  →  $*awi-$ ,  $*\mathfrak{ayi}-$  →  $*\etaawi-$ ), the original root turns out to be plausibly compatible with pre-PN  $*\eta\mathfrak{ay}-ti$ . Nara *keli* obviously does not belong here and is probably connected instead, through some old suffixal pattern, with *kela* ‘head, top’ q.v.

23. ‘**I**’. The basic form of the 1st p. sg. pronoun in most East Sudanic languages is  $*a-$ , usually extended with the suffixal component  $*-n-$  for the Southern groups and with  $*-k-$  for the Northern groups (cf. Bender’s division into “En” and “Ek” languages), although some variation does occur. The original variant is most clearly seen in Nara *a-g*; for Nubian  $*\mathfrak{a}-y$  (Mid. *áy*, ON *ay*, etc.) it is necessary to assume lenition of the velar stop, but the old root without the nominative singular marker is still preserved in some paradigmatic forms (e. g. Mid. accusative *á* ‘me’, etc.).

The biggest puzzle in this arrangement is PT *\*wa* 'I', reflected as such in most of the modern dialects. The appearance of an unpredictable *w*-, impossible to explain away as a prosthetic development or an enigmatic emphatic particle, makes the base pronominal paradigm of Tama incompatible on the whole with Nubian, Nara, and East Sudanic in general. On the other hand, even if one considers the typologically rare scenario of a borrowed origin for a basic personal pronoun, the fact remains that no modern areal neighbors of Tama have anything even remotely close to a *w*-shaped equivalent for 'I' — the closest would probably be Kanuri *wu*, but since there are no other reliable Tama-Kanuri or Tama-Saharan parallels in the basic lexicon, it is preferable to treat this phonetic similarity as coincidental.

One possible explanation comes from a comparison of this form with the paradigmatic peculiarities of the 1st p. pronoun in Hill Nubian, where it frequently takes on a labialized shape in the indirect stem (cf. Tagle *i*: 'I', gen. *ɔ-nná*, Dilling *e*, gen. *ɔ-ne*, etc.) and in Nara, where *a-g* 'I' is opposed to the genitive/dative stem (*w*)*o*. In light of this evidence, Claude Rilly has proposed to reconstruct a direct stem *\*a-(i)* and an indirect stem *\*o-* for Proto-Northeast Sudanic (Rilly 2009: 467), with analogical levelling in Proto-Tama (where languages such as Ibiri also show a separate genitive form *ho-n*). This does not quite explain why the nominative stem is *wa* and not the expected *\*o*, but the presence of these labialised indirect forms in Nubian and Nara is hardly accidental.

24. **'Kill'**. Nara *si-* and PT *\*siy-* (→ Mis. *siy-ɔ*, Ere. *ši-o*; Ibi. *ey*, Tama *iy-é* with regular loss of word-initial *s*-) present a perfect match. PN *\*pay-* (→ Nob. *fá:y-*, Mid. *pé-*, etc.) is not related and finds no clear correlates in the other two groups.

25. **'Leaf'**. Excluded from comparison. Most of the attested equivalents are either derived from the word for 'ear' (a very common typological development for the entire area) or are of obscure origin.

26. **'Louse'**. Nara *ši-ti* 'louse' has precisely the same phonetic shape as Ere., Sun. *ši-ti* id. (cf. also Tama *í-tì* with regular deletion of initial *\*s*-), although for PT, the original root shape has to be reconstructed as *\*sin-* based on Mis. *šin-ti* (the plural form is simply *šin*; special marking of the sg. rather than pl. number for this item is hardly surprising). Cluster simplification in Nara (*\*sin-ti* → *ši/t/ti*) is neither confirmed nor contradicted by additional examples, but is typologically plausible.

It is tempting to find some connection between these forms and PN *\*iti-di* 'louse' → Nob., Knz. *issi*, Dng. *issi*, Dil. *iti-d*, Mid. *i:di*, where *\*-di* is a relatively recent marker of the singulative, common in Nubian nominal stems. Theoretically, the remaining root *\*iti-* itself may be an old contraction from *\*iŋ-ti*, but there is no evidence that the initial sibilant could be deleted in PN just as it was (regularly) deleted in Tama; therefore, at the present stage the exact phonetic resemblance between such forms as Dilling *iti-* and Tama *í-tì* should rather be deemed a coincidence.

27. **'Meat'**. No parallels detected between any of the three taxa.

28. **'Moon'**. No parallels detected between any of the three taxa.

29. **'Mouth'**. The PN form is reconstructed rather securely as *\*agul-* (Knz., Dng. *agil*, Dil. *ɔgul*, Bir. *agal*, Mid. *á:l* with contraction; Nob. *ág* with seemingly regular deletion of stem-final *\*-l*). In Tama, the situation is more complicated: here, Eastern *\*kul* (Tama *kìl*, Ere., Sun., Mis. *kul*) seems poorly compatible with such Western forms as Ibi. *úli* ~ *awal*, AS *o:l* ~ *awl*. However, in Starostin 2014: 345 it was argued that both variants may still be reconciled under the assumption of two morphological variants in PT — simple *\*kul-* and its prefixal counterpart *\*V=kul-*, only the latter of which was preserved in the Western branch (with vocalic reduction and consonantal lenition: *\*V=kul-* → *\*awl-*). This solution remains hard to prove, but is nevertheless realistic (monovocalic fossilized prefixes were at least as likely to exist in PT as they were in PN), and makes the final reconstruction even more compatible with Nubian

data, since PT *\*V=kul-* may indeed have contained the very same prefix that is also preserved in PN *\*a<sup>1</sup>gul-*.

For Nara, it is essential to pay attention to the dialectal forms listed in Rilly 2009: 178: Higir *awlo*, Mogoreeb *àlkò*, Koyta *aulo*, Saantoorta *agura*. Although we do not have enough dialectal evidence to confirm this as part of a regular pattern, the only plausible way to explain the divergence is to set up the protoform *\*agulo* or *\*agula*, best preserved in Saantoorta (with a presumably regular *l* → *r* development) but undergoing reduction → *\*aglo* in the other dialects, with a subsequent metathesis in Mogoreeb and lenition → *\*aylo* → *\*awlo* in the other two dialects. This makes the form perfectly compatible with Nubian and Tama data.

30. ‘**Name**’. Nara *a:d-a* is obviously compatible with Proto-West Tama *\*a:t* (Ibi. *á:t*, AS *a:t*); Proto-East Tama *\*ŋa:t* (Ere., Mis. *ŋat*, Tama *ŋát*, Sun. *ŋat*) probably belongs here as well, provided that initial *ŋ=* may be viewed here as the same fossilized prefix that was already encountered above in ‘head’.

Rilly (2009: 486) suggests that both of these items are further compatible with PN *\*ə̥ri*, but there are too many unresolved problems with this comparison: even if the common Nara-Tama root is to be reconstructed as *\*a:d-*, there is no strong evidence that PNNT *\*-d-* could yield PN *\*-r-* in intervocalic position. Provisionally, we treat these etyma as different items.

31. ‘**New**’. No convincing parallels. Phonetic similarity is detected between PN *\*ə̥r* (Knz. *er*, Dng. *er*, Dil. *er*, Bir. *er*, etc.; replaced by a substrate element in ON *miri-*, Nob. *míri-*) and Nara *wor-ko* (Rn.), *wor-ku* (Bd.; also listed as *wer-* with a front vowel in Bender 1971: 268), but even if Nara *w-* is prothetic (of which there is no certainty), the significant difference in vocalism quality remains unexplained, so we provisionally reject this pair as a potential etymological match.

32. ‘**Night**’. A transparent isogloss between PN *\*a<sup>1</sup>war* (ON *oyar-*, Nob. *áwá*; Mid. *ò:d*; replaced in other branches by different innovations) and PT *\*war* (Tama *wàr*, Ere. *war*, Sun. *war-de*, Mis. *war*). In Nara, the old word was replaced by *kiše* ~ *kis-ne* (Rn.), *kisi-ŋa* (Bd.), bearing some resemblance to West Tama forms: Ibi. *išè*, AS *iše*. The latter, however, are transparent borrowings from nearby Maba (*išè* ‘night’), and since Maba-Nara contacts are geographically impossible, it is probably better to interpret the partial Nara – West Tama similarity as due to chance.

33. ‘**Nose**’. No parallels detected between any of the three taxa.

34. ‘**Not**’. All three taxa present evidence for at least two different morphemes that could mark indicative negation on the proto-level, but only one of them is compatible: PN *\*m-*, functioning as part of the negative verbal stem *\*mun- ~ \*min-* ‘not (to be)’ in Nile-Nubian and Hill Nubian and as a negative suffix in Birgid = Nara *ma* (negation marker in perfective forms) = PT *\*m-* (basic negative prefix in West Tama, also encountered as a prefix in certain adjectival stems in East Tama, cf. Sun. *áŋgé* ‘a lot’ vs. *m=áŋgē* ‘a little’). The others are different in all three taxa – PN monovocalic suffix *\*=a-*, fully preserved only in Midob but looking quite archaic in nature; Nara *ka* (negation marker in imperfective forms); and West Tama suffix *\*-to*. It is worth noting, however, that out of all East Sudanic languages, the only other family that shows signs of a proto-level *\*m*-shaped negative marker is Nilotc, so it is justified to regard this isogloss as significant.

35. ‘**One**’. No direct parallels detected between any of the three taxa. However, PN *\*bey- ~ \*bey-ir* ‘one’ (ON *we-l- ~ we-r-*, Nob. *wè: ~ wè:l ~ wè:-r*, Knz. *we:-r*, Dng. *we:-r*, Bir. *me:l-ug*, Mid. *pè:-r*; cf. also Mid. *pè:* ‘somebody’) is well compatible, phonetically and semantically, with Nara *bi-ko* (Rn.), *bi:k* (Bd.) ‘other’. Nara *doku* and PT *\*kV-* ‘one’ could only be related if *do-* in Nara were shown to be a prefixal component, which currently seems impossible.

36. ‘**Rain**’. PN *\*ar-* (ON *aru-*, Nob. *áwwí*, Dng. *aru*, Dil. *are*, Bir. *a:le*, Mid. *ár-*) is clearly the same root as East Tama *\*ar* (Tama *àṛ*, Ere., Mis. *àṛ*, Sun. *ar*). Whether Nara *hala* can belong here

as well is debatable: Rilly (2009: 501) lists the dialectal form *hàrà* from Saantoorta, but this seems to be the same dialectal development *\*l* → *r* as in ‘mouth’ (see above), and there are no other known cases of Nara *l* corresponding to PN and PT *\*r*. Initial *h*- also presents a problem; according to Rilly (2009: 302), it is an irregular reflex of PNNT *\*k*-, which seems to be well confirmed by several examples, so the overall correspondence for ‘rain’ in Nubian would be something like *\*kal-* rather than *\*ar-*.

37. ‘**Smoke**’. No direct parallels detected between any of the three taxa. However, it is permissible to compare PT *\*turu-* ‘smoke’ (preserved in Tama *túrú-t* and possibly in Ibi. *dùlód-à*, AS *dulud-a*, although correspondences are somewhat problematic) directly with Midob *tùrùd* ‘fog, mist’ (glossed this way in Werner 1993: 135, but mistakenly glossed as ‘smoke’ in Rilly 2009: 459).

38. ‘**Star**’. PN *\*wajn-* is best preserved in Birgid (*wajn-di*) and, with various contractions and assimilations, is also found in Hill Nubian (Kad. *wonɔ-ntu*, Deb. *won-du-nu*), Midob (*ònè-di*) and Nile-Nubian *\*wijn-di* ← *\*wajn-i-di* (ON *wijñ-*, Nob. *wìnzi*, Knz. *wissi*, Dng. *wissi*). All these forms are naturally compatible with Nara *wini* (Rn.) ‘star’ (Bender quotes the form *hū=wīnī*, where the first component is possibly the adjectival root ‘round’, cf. *hu-e* (Rn.) ‘to be round’).

More problematic is the relationship of these forms to PT *\*mijn-* ‘star’ (Tama *mijnì-t*, Ere. *minjì-t*, Sun. *mijn-a*; Ibi. *juñjì-t*, AS *jin-ti* with assimilation *\*m-* → *j-* due to the influence of the palatal nasal in word-medial position). On one hand, the most straightforward correspondence for this is the Nara verb *minni-* (Rn.) ‘to shine’. On the other hand, Tama data collected by Edgar shows a near-complete lack of native roots with the general structure *\*wVN-*, meaning that assimilation *\*wijn-* → *\*mijn-* would be perfectly natural in this protolanguage. Additionally, both Nara *wini* and PT *\*mijn-* display the same interesting polysemy ‘star/fly (n.)’ (not shared, however, by Nubian). In light of these observations, PT *\*mijn-* is judged as formally compatible with both PN and Nara and may be used as evidence for descent from the same common ancestral form (presumably *\*wajn-*, as in PN).

39. ‘**Stone**’. No direct parallels detected between any of the three taxa. ON *kit*, Nob. *kid* ‘stone’ are formally comparable with PT *\*kad-* (Mis. *kat*, Ere. *kadda*, Sun. *kada*), but the Nubian word is restricted to the Nobiin branch of Nile-Nubian, whereas the optimal distributional candidate for PN ‘stone’ is *\*kul-*, found in Kenuzi-Dongolawi, Birgid, and Midob; additionally, vocalic discrepancies are too severe here to make the Nobiin – Tama match a valid etymology.

40. ‘**Sun**’. No parallels detected between any of the three taxa.

41. ‘**Tail**’. No parallels detected between any of the three taxa.

42. ‘**Thou**’. The 2nd p. pronoun, unlike the 1st p., matches nicely across all three families, allowing to reconstruct *\*i-* as the simple root morpheme for PNNT (inherited from Common East Sudanic). In Nubian, *\*i-* shifts to *\*e-* in Birgid and in Kenuzi-Dongolawi (and then further to *\*a-* in Hill Nubian), but the original articulation is still well preserved in Nobiin and Midob. The oblique (genitive) stem *\*i-n-* is also common for PN and PT (Rilly 2009: 519).

43. ‘**Tongue**’. Our reconstruction *\*nalT-* for PN is significantly different from Bechhaus-Gerst’s *\*zardi*, but much closer to Rilly’s *\*ŋal*. The word-initial phoneme here is reflected as *\*n-* in Nile-Nubian (Nob. *nàr*, Knz. *ned*, Dng. *ned*), as *z-* in most Hill Nubian languages (Dil. *zal-e*, Kad. *zal-do*, Karko *zâr-è*, etc.), as *n-* in Birgid (*nat-ti*) and as *k-* (← *\*ŋ-*) in Midob (*kàd-i* ~ *kàd-àŋi*); Bechhaus-Gerst interprets it as *\*z-*, but this in no way explains the pervasive nasal reflexes. On the other hand, *\*ŋ-* is also excluded, since it is supposed to be preserved, not palatalized, in Hill Nubian. Based on the phonetic qualities of the different reflexes (coronal/velar nasals vs. palatal affricates), the optimal choice for reconstruction here is palatal *\*ŋ-*, and it seems to have been preserved in at least one Hill Nubian language: cf. Debri *ŋal-do* from Robin Thelwall’s field data (unless this is a misprint instead of *\*jaldo*).

Word-medially, we agree with Rilly that *\*-l-* rather than *\*-r-* should be reconstructed, since *\*-r-* is a highly stable phoneme in Nubian; however, a simple reconstruction of the root *\*jal-* (with a complex singulative correlate *\*jal-di*) does not suffice, since reflexes in individual languages are widely different from those of the similar stem *\*nil-di* ‘tooth’ (see below). Already in PN, the root itself must have contained a cluster (*\*jalt-* ~ *\*jald-*) or have been bisyllabic (*\*jalaT-*), which explains the loss of resonant articulation in Birgid (*nat-ti* ← *\*jalt-ti*) and word-final *-r* / *-d* in Nile Nubian (which usually appears in original *\*CVCV-ti* type structures, cf. ‘bone’ above).

This turns out to be significant on the level of external comparison, when the Nubian word for ‘tongue’ is compared with forms in Tama languages: Tama *àr(r)à-t*, Ere. *lā:t*, Sun. *lat*, Mis. *let*, Ibi. *lē:d* (also *la:t* and *laed* in alternate sources), AS *let*. This item is reconstructed as PT *\*la:t* by Rilly, but the reconstruction does not explain the front vowel in Mis. and Ibi., not to mention the odd diphthong *-ae-* in H. Barth’s and P. Doornbos’ transcriptions of West Tama material. In Starostin 2014: 360, it is argued that the discrepancies in vocalism and the diphthong-containing transcriptions can only be explained if *\*la:t* is traced back to an older *\*laCat*, where *\*-C-* is a weak consonant with palatalizing effect, most likely *\*-j-* (since glides like *-y-*, *-w-* do not regularly elide in intervocalic position).

The resulting reconstructions, PN *\*jalT-* (*\*julat-* ?) and PT *\*lajat*, are compatible under a simple metathesis scenario; the actual metathesis must have happened in Tama, as is indirectly hinted at by external data from other East Sudanic languages (cf. Nyimang *ŋildi*, etc.). Admittedly, this etymology rests rather heavily on intricacies of internal reconstructions in both Nubian and Tama, as well as upon assumption of irregular metathesis; however, irregularities and non-trivial developments are fairly typical of the word ‘tongue’ in numerous families all over the world. In any case, PN and PT are clearly more compatible with each other than Nara *haga*, an isolated form with no external parallels.

44. ‘Tooth’. A common feature of all three compared taxa is that they all share a nasal as the first consonant in the word for ‘tooth’: PN *\*jol-* (Knz. *nel*, Dng. *nel*, Bir. *jil-di*; Hill Nubian *\*ʒil-* → Dil. *ʒil-i*, Kad. *ʒil-du*, etc.; Mid. *kəd-dì* ← *\*jol-di*; Nob. *nì:d* ← *\*jil-d*), Nara *nihi*, PT *\*ŋes-* or *\*ŋeʒ-* (Ere. *niisi-t*, Sun. *niisi-t*, Mis. *ŋesi-t*; in Starostin 2014: 361, these forms are further compared with Ibi. *ŋónji-t*, AS *ŋoni-t* under a complex scenario of development from PT *\*ŋeʒ-*).

It seems, however, impossible to trace all three forms back to the same common source. There are two potential pathways here: (a) if the PT form is to be reconstructed as *\*ŋes-*, one could think of a common origin with PN *\*jol-*, showing the same hypothetical correspondence that had already been suggested earlier with ‘dog’ and ‘ear’, i.e. going back to PNNT *\*ŋet-*; (b) since Nara *nihi* must go back to *\*niKi* with an intervocalic velar stop, it might be compared with PT *\*ŋeʒ-* under the assumption of palatalization in PT (*\*negi* → *\*ŋeʒi*); unfortunately, there are currently no additional examples to support such an assumption. Curiously, external data from other East Sudanic languages provides evidence for *both* solutions: velar-medial forms are attested in Surmic (Southwest Surmic *\*nigi-t*, Southeast Surmic *\*nigi*), Jebel (*\*nigi*), and Daju (*\*niyi*) languages, whereas the lateral-medial form is seen in Nyimang (*\*ŋil-*; see tables in Starostin 2014: 722–729).

For the sake of uniformity, since we have not officially endorsed the correspondence of PN *\*l* to PT *\*s* yet, it is more prudent to go with the less radical variant (b) for the moment. Alternately, one could consider all three forms unrelated, but in the overall context of the situation, accidental similarity on all sides is hardly likely.

45. ‘Tree’. For PN, C. Rilly (2009: 423) reconstructs *\*ko:r-i* ‘tree’ vs. *\*ber-* ‘wood’; in Starostin 2014: 82, these reconstructions are amended to *\*koy/i/d* and *\*pər* respectively, and it is also pointed out that the latter word sometimes displays the polysemy ‘tree/wood’ (e. g. in

Hill Nubian or in old lexical materials on Kenuzi-Dongolawi) and should probably be projected in the meaning ‘tree’ onto the PN level, whereas the original meaning of *\*koy/i/d* may have been more narrow (e. g. = ‘Ziziphus spina-christi’ in Dng.). Recent innovation is also perceived in Nara, where Bender’s *kel* contrasts with *tûm* (Rilly’s spelling) ‘wood’, a word that is glossed as *tûm* ‘tree, wood’ in the old dictionary of Reinisch and is typologically likely to represent the older equivalent for ‘tree’. Even in Tama, the protoform *\*gan* ‘tree’ seems to be connected with the verbal root *ge- ~ gi-* ‘to rise, to stand up’ (diachronically, ‘to stand up, to be vertical’ is a well-known possible source for ‘tree’ as ‘vertically planted wood’, e. g. Chinese *shù*) and is probably secondary next to the old root *\*kij-* ‘wood’.

In any case, none of these forms match with each other, although some (Nara *tûm* and Tama *\*kij-*, in particular) may have interesting parallels in other branches of East Sudanic.

46. ‘**Two**’. Here, all the forms are compatible. In the case of Nubian, the most archaic form is found in Haraza Nubian *auri-yah* (Bell 1975: 84), which explains the non-trivial correspondence of Nile-Nubian *\*-ww-* (ON *uwo-*, Nob. *úwwó*, Knz. *owwi*, Dng. *owwi*) to Hill Nubian *\*-r-* (Dil. *ore-n*, Kad. *orrō*, Deb. *orrō*, Karko *ārē*). In Nara *ari-ga*, the labial element is missing (probably due to cluster simplification), but in PT *\*wari* (Tama *wári*, Ere. *wàrrí*, Sun. *warri*, Mis. *wora*, Ibi. *wári*, AS *werre*) it is found in word-initial position, suggesting metathesis: *\*awri* → *\*wari*.

47. ‘**Water**’. No direct parallels detected between any of the three taxa. PT *\*ka:l* (Tama, Ere. *ká:l*, Mis. *qal*, Ibi. *kàr-áŋ*, AS *kar-áŋ*) is etymologically comparable with Nara *kalli* (Rn.), *kalli* (Bd.) ‘cold’, since the semantic shift from ‘cold’ to ‘water’ is typologically plausible. External data from other East Sudanic languages suggest that Nara *mba* might be the most archaic form here (cf. Surmic *\*ma:m ~ \*maw*, Daju *\*ama ~ \*uma*, etc.), but comparable forms are not attested in either Nubian or Tama. The only possible exception is Old Nubian *aman-*, Nobiiin *ámán* ‘water, river, Nile’; however, distribution-wise this word belongs to the same layer of «Para-Nobiin substrate» as many other forms without Common Nubian etymologies, and cannot be reliably traced back to Proto-Nubian, let alone etymologically compared with Nara *mba*.

48. ‘**We**’. The PN reconstruction *\*a-y* is justified in detail in Starostin 2014: 86–90, where it is also argued that the clusivity opposition in certain Nubian languages (Midob, Old Nubian) is secondary and cannot be traced back to the PN level. It is quite tempting to put forward a plausible scenario in which PN *\*a-y* ‘T’ / *\*a-y* ‘we’ would directly correlate with Nara *a-g* ‘T’ / *a-gga* ‘we’ (e. g. PNNT *\*ag* → *\*ay* → *\*ay*, but PNNT *\*aga* → *\*aya* → *\*ay* without vocalic change), but it is hardly possible to back it with additional evidence. In any case, the pronouns here quite clearly match each other on the root level. As for Tama, *\*wa-yi* seems to be derived from *\*wa* ‘T’ (sg.), meaning that there are the same problems with trying to relate it to Nubian-Nara *\*a-* as with the singular correlate (see above).

49. ‘**What**’. In Nubian, there are two main groups of forms with the meaning of ‘what?': one beginning with *m-* (in Nile-Nubian: Old Nubian *mi-*, Nob. *mì-n*, Knz. *mi-n*, Dng. *mi-n-*) and one beginning with *n-* (Dil. *na*, Kad. *na-*, Bir. *na-ta*, Mid. *nè:-*, etc.). Rilly regards them as etymologically distinct, reconstructing *\*mi-n* and *\*na: ~ \*ne:* respectively. However, the second reconstruction is insecure, considering that the regular reflex of *\*n-* in Hill Nubian languages is *d-*, and in Midob it is *t-* (see ‘drink’ above). In Starostin 2014: 91, it is argued that the preservation of *\*n-* in this pronoun can only be due to some outstanding circumstances, and that under these circumstances the two forms may be traced back to a common protoform, provisionally given as *\*nWV-*, where *\*-W-* is an original labial glide or nasal. Such a form in itself could only be contracted from an earlier *\*nVwV-* or *\*nVmV-*, and this, in turn, makes it into an excellent match with Tama *nùmú-*, Ere. *numo-*, Sun. *nomo-*, Mis. *nùmá-*, Ibi. *nàma*, AS *nem-* ‘what’ ← PT *\*num*. Whether Nara *nda-* belongs here as well is far more debatable.

50. **‘Who’.** Nara *na* and PT \**na* (Tama *na-ye*, Sun., Mis. *na*, Ibi. *nà-n*, AS *na:-*) obviously match with each other. PN \**ŋɔ-y* is reconstructed with an initial velar nasal (this is most clearly seen in the Mid. reflex *kɔ:-*), which makes it hard to relate this root at least to PT \**na*, since initial \**ŋ-* is quite frequent in PT, and there are no obvious factors here that would explain the fronting \**ŋ- → \*n-* in PT. For now, we only count the Nara/Tama match as etymologically significant.

## Conclusions

Taking into consideration the importance of stratifying etymological and lexicostatistical matches to reflect their proportional representation across more and less stable layers of the basic lexicon, we separate the 50-item wordlist into a more stable and a less stable (on the average) half, based on the respective stability indexes of each item (see Table 1); Table 5 below summarizes the pairwise matchings in both halves found between all three taxa. Note that only the items that are marked with a + sign (i.e. credible etymological matches) in Table 1 are included in the calculations.

Table 5. Number of lexicostatistical matches between Nubian, Nara, and Tama.

|        | Nara       |             |         | Tama       |             |         |
|--------|------------|-------------|---------|------------|-------------|---------|
|        | Items 1–25 | Items 26–50 | Overall | Items 1–25 | Items 26–50 | Overall |
| Nubian | 8          | 5           | 13      | 8          | 10          | 18      |
| Nara   | —          |             |         | 10         | 5           | 15      |

The following conclusions may be drawn from the table itself, as well as from further analysis of some of the individual matches concealed behind the numbers.

1. The highest number of matches is between PN and PT:  $18/50 = 36\%$ . This is much higher than the 20% figure given in Starostin 2014: 677, where only the automatically detected pseudo-cognates were counted. However, both of these figures are statistically significant (based on empiric evidence from comparing multiple random pairs of unrelated languages, we accept a threshold of 5-6 matches out of 50 to rule out accidental similarity), and the same is true for the other two pairs as well<sup>11</sup>.

2. Using Indo-European as a comparative benchmark, we may select, e.g., Old Indian as the approximate chronological equivalent of PN, and Proto-Germanic or Latin as the approximate chronological equivalent for the somewhat younger PT. In this case, the figure of 36% will be significantly lower than the corresponding numbers for Sanskrit vs. Latin (57%) or Sanskrit vs. Germanic (56%)<sup>12</sup>. This means that if Nubian and Tama languages are genetically related, their common ancestor must have probably been older than Proto-Indo-European (e.g., Sergei Starostin’s recalibrated glottochronological formula in this case yields a dating of approximately 4700 BC).

<sup>11</sup> Had this number of parallels been seriously lower (e.g. in the range of 8–10 matches out of 50), it would have made sense to apply the same kind of permutation test as performed in, e.g., Kassian, Zhivlov, Starostin 2015, in order to establish statistical significance on a formally rigorous basis. With this amount of evidence, however, it hardly seems worth the bother.

<sup>12</sup> These numbers are based on preliminary 50-item wordlists, reconstructed or collected for various small language groups of Eurasia and publicly available on the Global Lexicostatistical Database website: <http://starling.rinet.ru/new100/eurasia.xls>.

3. The overall numeric correlations between Nubian, Nara, and Tama give no definitive answer to the question of the internal structure of their phylogenetic tree. Although 18 matches between Nubian and Tama is a significantly higher number than 13 matches between Nubian and Nara, this is primarily explicable by the fact that Nara is a modern language, while PT is a reconstruction that pushes us back about 2000 years, so that, even if all three branches split from their common source at the same time, we would naturally expect Nara to show less in common with PN and PT than both of them have in common with each other. At the moment, all three taxa appear to be more or less equidistant; future studies will let us understand better if there are any truly decisive shared innovations in between any two out of three branches of the family.

4. The distribution of cognates across the various stability groups correlates very well with our expectations (more cognates in the more stable part, fewer cognates in the less stable part) in the case of Nubian-Nara (8 against 5) and Nara-Tama (10 against 5), but not in the case of Nubian-Tama (8 against 10) — due to such shared items as ‘ashes’, ‘egg’, ‘head’, ‘horn’, ‘night’, ‘rain’ that have no parallels in Nara. Although the discrepancy is not altogether tragic, it does suggest that at least a few of these matches might ultimately be areal rather than genetic in origin: for instance, the word *\*ar-* for ‘rain / sky’ has a rather wide areal distribution and could represent a cultural *Wanderwort* rather than an inherited term.

5. On the other hand, it is notable that cognates are encountered across all semantic and functional classes of words — including body part terms, verbs, personal and interrogative pronouns, and even the negation marker. Combined with additional etymologies and occasional grammatical isoglosses that were previously published in J. Greenberg's, M. L. Bender's, and C. Rilly's works, this makes the scenario of common descent from a Proto-Nubian-Nara-Tama ancestor far more plausible than the opposite scenario of areal diffusion.

It must be stressed that, although the absolute majority of lexical parallels commented upon in this paper had previously been suggested by at least one or more of the abovementioned authors, the sort of etymological / lexicostatistical refining conducted here — where only direct semantic matches are taken into consideration, and each candidate for comparison is vetted on the issue of reconstructibility for proto-status, to reduce the risk of accidental matches — has been performed for the first time. In our opinion, the Nubian-Nara-Tama connection passes this restrictive test with flying colors. On the other hand, the question of whether it makes practical sense to try to produce a large etymological corpus solely for PNNT without taking into consideration the data from other East Sudanic languages is still open: as we have seen, NNT is a fairly deep family, probably older than Indo-European by at least one millennium, and this, combined with the relative scarceness of data on Nara and Tama as well as several millennia of areal interference, means that positive identification of large numbers of cognates is going to be a very hard task without assessing the hypothesis in an even larger context. The next logical step for such an assessment would be to investigate the position of Nyimang, a minor language group of Kordofan whose ties to NNT seem to be counterbalanced with its ties to the neighboring Temein languages; we plan to cover this issue in our next publication on East Sudanic lexicostatistics.

#### Abbreviations

AS — Abu Sharib; Bd. — Bender 1968; Bir. — Birgid; Deb. — Debri; Dil. — Dilling; Dng. — Dongolawi; Ere. — Erenga; Ibi. — Ibiri; Kad. — Kadaru; Knz. — Kenuzi; Krk. — Karko; Mid. — Midob; Mis. — Miisiirii; Nob. — Nobin; ON — Old Nubian; PN — Proto-Nubian; PNNT — Proto-Nubian-Nara-Tama; PT — Proto-Tama; Rn. — Reinisch 1874; Sun. — Sungor; Wal. — Wali.

## References

Abushush, Dawd, Richard J. Hayward. 2002. Nara. *Journal of the International Phonetic Association* 32(2): 249–255.

Armbruster, Charles H. 1965. *Dongolese Nubian: A Lexicon*. Cambridge University Press.

Bechhaus-Gerst, Marianne. 1985. Sprachliche und historische Rekonstruktionen im Bereich des Nubischen unter besonderer Berücksichtigung des Nilnubischen. *Sprache und Geschichte in Afrika* 6(1984/5): 7–134.

Bechhaus-Gerst, Marianne. 1989. “Nile-Nubian” reconsidered. In: M. Lionel Bender (ed.) *Topics in Nilo-Saharan Linguistics*: 85–96. Hamburg: Helmut Buske Verlag.

Bechhaus-Gerst, Marianne. 1996. *Sprachwandel durch Sprachkontakt am Beispiel des Nubischen im Niltal: Möglichkeiten und Grenzen einer diachronen Soziolinguistik*. Köln: Rüdiger Köpfe Verlag.

Bechhaus-Gerst, Marianne. 2011. *The (Hi)story of Nobiin — 1000 Years of Language Change*. Frankfurt am Main: Peter Lang.

Bell, Herman. 1970. The Phonology of Nobiin Nubian. *African Language Review* 9: 115–139.

Bell, Herman. 1975. Documentary evidence on the Haraza Nubian language. *Sudan Notes And Records* 56: 1–35.

Bender, Lionel M. 1968. Analysis of a Barya wordlist. *Anthropological Linguistics* 10(9): 1–24.

Bender, Lionel M. 1971. The Languages of Ethiopia: A New Lexicostatistic Classification and Some Problems of Diffusion. *Anthropological Linguistics* 13(5): 165–288.

Bender, Lionel M. 1997. *The Nilo-Saharan Languages: A Comparative Essay*. München – Newcastle: LINCOM Europa.

Bender, Lionel M. 2005. *The East Sudanic Languages: Lexicon and Phonology*. Southern Illinois University: SIU Printing/Duplicating.

Blench, Roger. 2011. *Can Sino-Tibetan and Austroasiatic help us understand the evolution of Niger-Congo noun classes?* Ms. (presentation given at the 41st CALL, Leiden, August 29–31, 2011).

Browne, Gerald M. 1996. *Old Nubian Dictionary*. Leuven: Peeters.

Browne, Gerald M. 2002. *A Grammar of Old Nubian*. Münich: LINCOM.

Bryan, Margaret A. 1959. The T-K Languages. A New Substratum. *Africa: Journal of the International African Institute* 29(1): 1–21.

Burlak, S. A., S. A. Starostin. 2005. *Sravnitel'no-istoricheskoe jazykoznanije [Comparative-Historical Linguistics]*. Moscow: Academia.

Edgar, John T. 1990. *Tama Group Lexicon*. Ms.

Edgar, John T. 1991a. First Steps Toward Proto-Tama. In: M. Lionel Bender (ed.). *Proceedings of the 4th Nilo-Saharan Conference. Bayreuth, Aug. 30 – Sep. 2, 1989*: 111–131. Hamburg: Helmut Buske Verlag.

Edgar, John T. 1991b. *Maba-Group Lexicon*. Berlin: Dietrich Reimer Verlag.

Ehret, Christopher. 2001. *A Historical-Comparative Reconstruction of Nilo-Saharan*. Köln: Rüdiger Köpfe Verlag.

Fleming, Harold. 1983. Kuliak External Relationships: Step One. In: R. Voßen & M. Bechhaus-Gerst (eds.). *Nilotic Studies*: 423–478. Berlin: Dietrich Reimer Verlag.

Greenberg, Joseph H. 1966. *The Languages of Africa*. Bloomington, Indiana University; Mouton & Co., The Hague.

Hammarström, Harald, Robert Forkel, Martin Haspelmath. 2017. *Glottolog 3.0*. Jena: Max Planck Institute for the Science of Human History. Available online at <http://glottolog.org>; accessed on 2017-05-21.

Hayward, Richard J. 2000. Observations on Tone in the Higir Dialect of Nara. In: R. Vossen, A. Mietzner, A. Meissner (eds.). “*Mehr als nur Worte...*”: *Afrikanistische Beiträge zum 65. Geburtstag von Franz Rottland*: 247–267. Köln: Rüdiger Köpfe Verlag.

Heusing, Gerald. 2004. *Die südlichen Lwoo-Sprachen. Beschreibung, Vergleich und Rekonstruktion*. Köln: Rüdiger Köpfe Verlag.

Hofmann, Inge. 1986. *Nubisches Wörterverzeichnis: Nubisch-deutsches und deutsch-nubisches Wörterverzeichnis nach dem Kenzi-Material des Samuël Alí Hisén (1863–1927)*. Berlin: Dietrich Reimer Verlag.

Jabr el Dar, Khaliifa. 2006. Towards a general orthography of the Ajang languages. In: Al-Amin Abu-Manga, Leoma Gilley, & Anne Storch (eds.). *Insights into Nilo-Saharan Language, History and Culture*: 183–198. Köln: Rüdiger Köpfe Verlag.

Jakobi, Angelika. 2006. The loss of syllable-final proto-Nubian consonants. In: Al-Amin Abu-Manga, Leoma Gilley, & Anne Storch (eds.). *Insights into Nilo-Saharan Language, History and Culture*: 215–228. Köln: Rüdiger Köpfe Verlag.

Kassian, Alexei. 2015. Towards a Formal Genealogical Classification of the Lezgian Languages (North Caucasus): Testing Various Phylogenetic Methods on Lexical Data. *PLoS ONE* 10(2): e0116950 (<https://doi.org/10.1371/journal.pone.0116950>).

Kassian, Alexei, George Starostin, Anna Dybo, Vasily Chernov. 2010. The Swadesh wordlist: an attempt at semantic specification. *Journal of Language Relationship* 4: 46–89.

Kassian, Alexei, Mikhail Zhivlov, George Starostin. 2015. Proto-Indo-European-Uralic Comparison from the Probabilistic Point of View. *Journal of Indo-European Studies* 43: 60–80.

Kauczor, P. D. 1920. *Die Bergnubische Sprache (Dialekt von Gebel Delen)*. Wien: Alfred Hölder.

Khalil, Mokhtar M. 1996. *Wörterbuch der nubischen Sprache (Fadidja/Mahas Dialekt)*. Warszawa: Piotr O. Scholz.

Krell, Amy. 2012. *Rapid Appraisal Sociolinguistic Survey Among Ama, Karko, and Wali Language Groups (Southern Kordofan, Sudan)*. SIL International.

Lepsius, Carl Richard. 1880. *Nubische Grammatik. Mit einer Einleitung über die Völker und Sprachen Afrikas*. Berlin: W. Hertz.

Lukas, Johannes. 1933. Beiträge zur Kenntnis der Sprachen von Wadai. *Journal de la Société des Africanistes* 3(1): 25–55.

Lukas, Johannes. 1938. Die Sprache der Sungor in Wadai (Aus Nachtigals Nachlaß). *Ausland Hochschule Mitteilungen* 41(3): 171–246.

MacMichael, H. A. 1920. Darfur Linguistics. *Sudan Notes and Records* 3(3): 197–216.

Massenbach, Gertrud von. 1962. *Nubische Texte im Dialekt der Kunuzi und der Dongolawi, mit Glossar*. Wiesbaden: Deutsche Morgenländische Gesellschaft.

Meinhof, Carl. 1918. Sprachstudien im ägyptischen Sudan. 33. Dulman. 34. Garko. 36. Kadero. 37. Koldegi. 38. Zur Formenlehre der nubischen Dialekte. 39. Vergleichendes Wörterverzeichnis: Deutsch-Nubisch. 40. Kenuzi. 41. Dongola. 42. Fadidja. 43. Mahas. 44. Bedauya. *Zeitschrift für Kolonialsprachen* IX: 43–64, 89–117, 167–204, 226–255.

Reinisch, Leo. 1874. *Die Barea-Sprache. Grammatik, Text und Wörterbuch. Nach den handschriftlichen Materialien von Werner Munzinger Pascha*. Wien: Wilhelm Braumüller.

Rilly, Claude. 2005. The Classification of Nara Language. *Journal of Eritrean Studies* 4: 1–27.

Rilly, Claude. 2009. *Le Méroïtique et sa famille linguistique*. Louvain – Paris – Dudley, MA: Peeters.

Starostin, George. 2010. Preliminary lexicostatistics as a basis for language classification: a new approach. *Journal of Language Relationship* 3: 79–117.

Starostin, George. 2016. From wordlists to proto-wordlists: reconstruction as ‘optimal selection’. In: K. Pozdniakov (ed.). *Comparativism and reconstruction: tendances actuelles. Faits de Langues* 47: 177–200. Berne: Peter Lang.

Starostin, Georgij. 2013. *Jazyki Afriki. Opyt postrojenija leksikostatisticheskoy klassifikacii. Tom I: Metodologija. Khojanskije jazyki [Languages of Africa. An attempt at a lexicostatistical classification. Volume I: Methodology. Khoisan languages]*. Moscow: Jazyki slav'anskoy kul'tury.

Starostin, Georgij. 2014. *Jazyki Afriki. Opyt postrojenija leksikostatisticheskoy klassifikacii. Tom II: Vostochnosudanskije jazyki [Languages of Africa. An attempt at a lexicostatistical classification. Volume II: East Sudanic languages]*. Moscow: Jazyki slav'anskoy kul'tury.

Starostin, Sergei. 2000. Comparative-historical Linguistics and Lexicostatistics. In: C. Renfrew, A. MacMahon, L. Trask (eds.). *Time Depth In Historical Linguistics*: 223–259. McDonald Institute for Archaeological Research, Oxford Publishing Press.

Starostin, Sergei. 2007. Opredelenije ustojčivosti bazisnoj leksiki [Defining the Stability of Basic Lexicon]. In: Sergei Starostin. *Trudy po jazykoznaniju [Works in Linguistics]*: 825–839. Moscow: Jazyki slav'anskix kul'tur.

Thelwall, Robin. 1977. A Birgid vocabulary list and its links with Daju. In: H. Ganslmayr & H. Jungraithmayr (eds.). *Gedenkschrift Gustav Nachtigall 1874–1974*: 197–210. Bremen: Übersee-Museum.

Thelwall, Robin. 1978. Lexicostatistical relations between Nubian, Daju and Dinka. In: J. Leclant, J. Vercoutter (eds.). *Etudes Nubiennes. Colloque de Chantilly, 2–6 Juillet 1975*: 265–286. Caire: L’Institut Français d’Archéologie Orientale du Caire.

Thelwall, Robin. 1983. Meidob Nubian: Phonology, Grammatical Notes and Basic Vocabulary. In: Lionel M. Bender (ed.). *Nilo-Saharan Language Studies*: 97–113. Michigan: East Lansing.

Thompson, E. David. 1976. Nera. In: M. Lionel Bender (ed.). *The Non-Semitic Languages of Ethiopia*: 484–494. East Lansing: Michigan State University.

Tucker, Archibald, Margaret A. Bryan. 1956. *The Non-Bantu Languages of North-Eastern Africa*. Oxford University Press.

Vasilyev, Mikhail, George Starostin. 2014. Leksikostatisticheskaja klassifikacija nubijskikh jazykov: k voprosu o nil'sko-nubijskoj jazykovoj obshnosti [Lexicostatistical classification of the Nubian languages and the issue of the Nile-Nubian genetic unity]. *Journal of Language Relationship* 12: 51–72.

Vossen, Rainer. 1982. *The Eastern Nilotes. Linguistic and historical reconstructions*. Berlin: Dietrich Reimer Verlag.

Werner, Roland. 1987. *Grammatik des Nobiin (Nilnubisch). Phonologie, Tonologie und Morphologie*. Hamburg: Helmut Buske Verlag.

Werner, Roland. 1993. *Tidn-áal: A Study of Midob (Darfur Nubian)*. Berlin: Dietrich Reimer Verlag.

Youn, Hyejin, Logan Sutton, Eric Smith, Christopher Moore, Jon F. Wilkins, Ian Maddieson, William Croft, Tanmoy Bhattacharya. 2016. On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences of the United States of America* 113(7): 1766–1771.

Zyhlarz, Ernst. 1950. Die Lautverschiebung des Nubischen. *Zeitschrift für Eingeborene Sprachen* 35: 1–20, 128–146, 280–313.

Г. С. Старостин. Лексикостатистические исследования по восточносуданским языкам I: к вопросу о нубийско-нара-тама генетической общности

В статье дается подробный лексикостатистический обзор реконструированных 50-словных списков (сокращенный вариант классического списка Сводеша, состоящий из более устойчивых элементов) по трем языковым группам северо-восточной Африки — нубийской, нара и тама. Эти группы традиционно относятся к восточносуданской семье и в большинстве существующих классификаций описываются как особенно близко родственные друг другу. В обзоре продемонстрировано, что как в количественном, так и в качественном отношении лексикостатистические параллели между нубийскими, нара и тама языками убедительно интерпретируются как следы общего происхождения (а не ареальной близости) этих групп, что формально подтверждает гипотезу, которой придерживались Дж. Гринберг, М. Л. Бендер, К. Рильи и другие исследователи. При этом глоттохронологическая оценка гипотезы показывает, что пранубийско-нара-тама язык следует относить к периоду не позднее 5-го тыс. до н. э., т. е. семья в целом оказывается даже более древней, чем праиндоевропейская, и насколько детально можно будет реконструировать для нее этимологический корпус, остается неясным. Статья представляет собой первую публикацию из серии, которую предполагается посвятить комплексной этимолого-лексикостатистической оценке восточносуданской гипотезы.

**Ключевые слова:** нило-сахарские языки, восточносуданские языки, нубийские языки, языки тама, африканское историческое языкознание.