George Starostin
Russian State University for the Humanities / RANEPA (Moscow)

Macro-comparative linguistics in the 21st century: state of the art and perspectives

The paper represents an attempt to explicitly summarize most of the major theoretical and methodological problems that, as of today, hinder significant progress in the field of macro-comparative linguistics (research on distant relationships between the various language families of the world). Among these problems are such issues as: the amount, quality, and nature of linguistic data that is necessary to establish long-distance relationship; methodological priorities of the etymologization process; and the complex interdependencies of "objective" (automated) and "subjective" (manual) data comparison. Partial solutions and/or recommendations of a general character are offered for each of the specified issues.

Keywords: Comparative linguistics, long-range comparison, language macrofamilies, computational methods in linguistics.

Any attempt to exhaustively summarize the state of the art of modern macro-comparative linguistic studies within the limits of a short paper would result in inevitable failure. Not even the most informed linguist these days could lay claim to being fully aware of everything that is going on in this field. As a matter of fact, no responsible linguist would probably *want* to waste precious time on such completism, since, in many respects, the field in question is actually a "minefield", in that the reasonably accurate long-ranger, attempting to navigate it with proper scientific diligence and caution, does not always understand where the principal opposition is coming from. Should the long-ranger, first and foremost, engage in theoretical and methodological arguments with "traditionally-minded" historical linguists? With philologists? With computational scientists? Or should he/she be primarily concerned about rectifying the methods and hypotheses of his/her other fellow long-rangers, whose research has occasionally made macro-comparison so easily vulnerable to criticism from all the abovementioned groups of people?

With this question in mind, I would like to use this short paper as a pretext to briefly discuss where we stand today in the long-range comparison department from a *theoretical* point of view, rather than (as the title could suggest) simply list a bunch of macrofamily hypotheses and rank them in terms of how much credit they have (if any at all) among mainstream specialists[1]. (We will refrain from the delicate discussion of what is actually supposed to constitute the "mainstream" paradigm in today's historical linguistics — for the purposes of the current paper, it will suffice to equate "mainstream specialists" with the average anonymous peer reviewer who is automatically tempted to give a negative assessment to any paper that dares to mention J. Greenberg or V. M. Illich-Svitych in a positive light).

It is more or less obvious that there is little sense to talk about how "reliable" or "convincing" a particularly bold hypothesis of genetic relationship, such as Nostratic or Amerind,

[1] For a summary of these hypotheses from the point of view of the "Moscow school of comparative linguistics" and the Evolution of Human Languages project (http://ehl.santafe.edu), see the detailed, if already slightly obsolete, report, published several years ago as Gell-Mann, Peiros, Starostin 2009.

is for a particular researcher, unless *all* researchers, or, at least, a sizeable and representative group of researchers, has reached a detailed agreement on the *criteria* according to which the hypothesis should be judged. Actually, it is not even the word “criteria”, but rather the word “*detailed*” that should be stressed — time has shown, over and over again, that several people can easily agree on a general principle (such as “regular phonetic correspondences”, or “semantic proximity”, or even “lexicostatistical analysis”), but when it comes to practical research, it often turns out that everyone has one’s own understanding of what these words are supposed to mean in particular situations.

The result is a huge conflict of interest even between the long-rangers themselves. Whenever a macro-hypothesis is being worked on by several people that do not form a coherent team, what one gets is a sub-set of highly dissimilar and often mutually incompatible sub-hypotheses: not “Nostratic” as such, but “Illich-Svitych’s Nostratic”, “Dolgopolsky’s Nostratic”, “Bomhard’s Nostratic”[2]; not “Nilo-Saharan”, but “Greenberg’s Nilo-Saharan”, “Bender’s Nilo-Saharan”, “Ehret’s Nilo-Saharan”[3], etc. This confusion turns both the hypotheses and their authors into easy prey for critics of long-range comparison, who are under no obligation to conduct a proper comparison of all the competing variants — even if they may sometimes be mildly sympathetic to the more accurately stated ones, they still run the risk of being discarded along with the poor ones (for extra security).

But there is an additional problem here, the gravity of which, I believe, is not always fully understood either by the “lumpers” or the “splitters”. Today, it is no longer possible for mainstream linguistics to simply *ignore* the macro-comparative department or to *dismiss* it right out of hand, either explicitly or implicitly[4]. The main reason for this is increasing pressure from the adjoining disciplines — above all, archaeology and human genetics.

Modern genetics, in particular, since it studies human prehistory from a natural science angle, is employing sets of powerful computational tools to construct credible genetic trees of humanity and trace its major migration routes — while it isn’t quite there yet, it is highly likely that only a matter of years, perhaps a few decades, separates us from having at our disposal a mathematically approved “optimal” scenario of the chronological dispersal of *Homo sapiens*. And, naturally, geneticists are all too keen to compare the results that they are getting — for periods that go back dozens of thousands of years — with whatever the linguists have to offer; and, on an informal basis at least, they are frequently quite surprised to discover that the linguists do not have to offer *that* much, even though they have been going at it for a far longer period of time.

In all seriousness, it no longer becomes permissible for the comparative linguist to remain content with hiding behind an agnostic formula, such as “the comparative method cannot go further back than six…” (eight, ten, the precise figure does not matter) “…thousand years”[5]. It is becoming a matter of “do or die”: the comparative method *has* to be able to go further than

[2] The three competing variants of the Nostratic theory, supported with intersecting, but still widely varying evidence, are best illustrated in their authors’ etymological dictionaries: Illich-Svitych 1971–1984, Dolgopolsky 2008, Bomhard 2003.

[3] See Greenberg 1966, Bender 1997, Ehret 2001 for three entirely different methodologies and vastly differing corpora of evidence to back up the hypothesis.

[4] “Implicit” rejection may, for instance, be found in a recent monograph on language classification (Campbell & Poser 2008), where the authors express mild optimism about the future of demonstrating “remote linguistic relationships”, but it soon turns out that their ideal of a remote linguistic relationship is represented by such families as Uralic or Uto-Aztecan, which no authentic long-ranger could ever define as “remote”.

[5] Cf.: “the diagnostic kinds of accidence… dissipate entirely after about 8000 years (distinctly less in languages of certain non-Indo-European structural types)” (Nichols 1992: 313).

that, and even if someone manages to prove beyond reasonable doubt that it cannot (which, in my opinion, is not very likely), in reality this will only mean that the *current version* of the comparative method is not robust enough to go there — so it is up to us to increase its robustness and succeed where previous efforts have failed.

Giving up on the comparative method when "larger-than-*x*" time periods are concerned means only one thing: something else will inevitably emerge to take its place in linguistic palaeontology, and that "something else" will not necessarily be an improvement. In fact, something else has *already* emerged: the application of computational phylogenetic models, borrowed from biology and other branches of natural science, to linguistic data — a development that has enjoyed ever increasing popularity throughout the last decade ever since Russell Gray's and Quentin Atkinson's (2003) landmark publication of their model of Indo-European classification in "Nature". The so-called "Gray Lab" in Oakland, New Zealand still remains arguably the largest single supplier of such studies, but multiple other teams and individuals have also joined the fray, and while, on the whole, these developments are quite exciting, one generally troubling thing about them is that most of the involved scientists are anything *but* professionally trained historical linguists. Paradoxical as it may sound, the most widely discussed publications in historical linguistics from the last decade have not been published by actual historical linguists! And this fact alone should be enough to make most historical linguists — including "long-rangers" / "lumpers" and "short-rangers" / "splitters" alike — sit up and take notice; especially because, in this particular case, "the most widely discussed" may not necessarily mean "the most useful in advancing real science".

There is, however, one aspect in all of these recent computational studies that "traditional" historical linguists should probably pay close attention to — all of them approach linguistic reconstruction and classification as *stochastic* in nature, usually generating results by exploiting various models that are based on the principles of maximum likelihood. And while some might object that the stochastic principle in comparative linguistics may serve as a clever trick to make the work formally invulnerable to criticism ("why does this model contradict established historical facts?" — "well, the model does not state that the results are *right*, it just states that this is the most probable outcome given the algorithm and the data fed into the algorithm"), probabilistic reasoning is actually a deeply right methodological foundation that is all too often overlooked in traditional comparative linguistics — namely, that the process of linguistic reconstruction is almost always a process of *choice* of an optimal solution given the circumstances, rather than a process of strict mathematical or logical *proving* of any particular statement. Despite the frequent use of the very loosely understood term "proof" in comparative linguistics, one does not usually resort to rigorously proving anything in this discipline; instead, one considers alternatives, for each of which evidence is gathered, weighed, and the weightier bunch is given precedence. It is perfectly possible that tomorrow new evidence will appear, forcing us to switch from solution A to solution B, making the latter "weightier" than it was yesterday — but this, of course, is just how science is actually expected to work.

So where does today's macro-comparative linguistics stay on this issue? Let us roughly assume the existence of two principal schools of long-range comparison. One of them insists on the helplessness of the comparative method when it comes to sorting out superfamilies — this is the "mass comparison" approach of Joseph Greenberg and his followers. The other one is the tradition initiated by Russian linguists Vladislav Illich-Svitych and Aharon Dolgopolsky with their "Nostratic" hypothesis and subsequently inherited by the so-called "Moscow school" of comparative linguistics: its main idea is that there are no fundamental differences between long-range comparison and regular short-range comparison of the Neogrammarian type. One look at the voluminous Nostratic dictionary by A. Dolgopolsky (2008) clearly suffices to reveal the stylistic legacy of J. Pokorny's Indo-European dictionary (somewhat obscured as it is by a swarm of additional transcriptional symbols and complex abbreviations). In fact, both schools agree in their assumption that the same methodology, be it "mass comparison" or the classic comparative method, may and should be applied to language families of just about any time depth.

The relations of both approaches to the issue of "probability" and "optimality" is complex. Greenberg, harassed by demands of "rigorous proof" by traditionally-oriented historical linguists throughout his life, recognized and advocated the importance of probabilistic thinking in almost every theoretical publication where he defended mass comparison[6]. Unfortunately, it might be argued that he hardly ever practiced what he preached: in his huge corpora of data, Greenberg rarely discusses alternatives, never really drawing the reader's attention to what it is exactly that makes *his* hypotheses better than anybody else's, not to mention the "null hypothesis" of non-relationship. Hence the superficially convincing "discreditation" of the Amerind hypothesis by Lyle Campbell (1988), who successfully "demonstrated" that Finnish was Amerind, too; all the more easily done since there was no strict procedure to quantify the evidence in Greenberg's paradigm — no way to state explicitly how the evidence for "Finnish as Amerind" relates in quantitative or probabilistic terms to the evidence for "Amerind as Amerind".

On the other hand, the tradition of regarding macro-comparative studies as simply an extension of regular comparative studies, upheld most rigorously in the Moscow school, still shares the same methodological problem as its opponents: some scholars still seem too beset with the idea of "categoricity", the issue of "proof" of relationship understood in strictly binary terms (yes or no) with a dogmatic flair. Pro- and opponents of long-range comparison alike demand a threshold for the evidence, one which they rarely, if ever, try to establish formally and, therefore, one which they all understand differently. For instance, the demands that Lyle Campbell establishes for "proving" a long-range comparison are so strict that, most likely, not a single genuine long-range hypothesis will ever be proven to his satisfaction (Starostin 2009). At the same time, one could also list certain long-rangers, even those who formally accept the importance of the comparative method and regular phonetic correspondences, whose own threshold is intuitively motivated by a small handful of intuitively stunning resemblances between families, around which they build up pseudo-theories that are completely devoid of historical realism.

All of my own experience of working on long-range comparison, mostly in the area of large Eurasian stocks ("Nostratic", "Sino-Caucasian") and African classification, seems to indicate the necessity of admitting it — there *is*, or at least, there *should* be a certain methodological difference between the usual ways in which comparative linguistics has up to now worked with "short-range" families, and the ways in which we should be conducting further research on the macro-comparative level. Furthermore, I believe that the fact that macro-comparative linguistics has somewhat stalled these days, leading to a temporary crisis of such studies, has much to do with our reluctance to accept this.

Hence, if we want real progress in the development of macro-comparative studies, it makes sense to suggest that a certain compromise is necessary between all these positions, one that would take the most promising aspects of "macro-comparative Neogrammarianism" *à la* Illich-Svitych / Dolgopolsky, of Greenberg's "mass comparison", of modern computational

[6] E. g.: "…in all emprical sciences… all that we *can* get are results so close to certainty that for all practical purposes we can consider them true, that is, a hypothesis which is overwhelmingly better than any other in accounting for the facts" (Greenberg 1995: 207).

approaches, assess them in the light of reasonable, unbiased "traditionalist" criticism, and synthesize a relatively complex, but "accessible" and efficient set of methodological rules. Such a thing is probably easier said than done, as it clearly involves a lot of heavy, often tedious teamwork, where specialists working on different language families should regularly share their experience and help refine each other's work ideology; but the more tedious that teamwork is, the less space might be left for creative bickering between people who should be advancing science rather than wasting time on highly unproductive debate (as is, unfortunately, quite often the case in historical linguistics).

In an attempt to help facilitate such a compromise, I will try to concisely outline four of the most common points of debate that are regularly observed in works on long-range comparison (actually, all four are relevant for historical linguistics as a whole, but the relevance naturally increases as we proceed further on to deeper time levels).

Problem 1: Quantity or quality?

This dilemma, as a rule, tends to be resolved in extremities. Active and productive proponents of macro-comparative studies tend to stress the importance of compiling huge etymological dictionaries (cf. approximately 2800 etymologies in the comparative Altaic dictionary of Starostin, Dybo, and Mudrak (2003); amazingly, almost exactly the same number — 2800! — in Dolgopolsky's Nostratic dictionary (2008); more than 3000 etymologies in the Afro-Asiatic database by A. Militarev and O. Stolbova (available online at http://starling.rinet.ru); 1606 etymologies in Ch. Ehret's Nilo-Saharan dictionary (2001), etc.). A common criticism of such "etymological mastodonts" is that these huge numbers are only possible since there are so many languages to choose from — implying that most, if not all, of the individual etymologies simply reflect chance resemblances that accumulate in daughter languages as time passes.

The opposite approach is illustrated by such relatively recent attempts at macro-comparison as Laurent Sagart's "Sino-Austronesian" (2005) or Edward Vajda's "Dene-Yeniseian" (2011). These works tend to focus on a relatively small number of examples (around 100 etymologies); to introduce a system of phonetic correspondences that is claimed to be fully regular (something that Eric Hamp (2011) calls "total accountability"); and to exclude any semantic comparisons that could seem improbable or dubious. Such comparisons have a lot of potential charm — above all else, they are particularly seductive for the eye of the outside evaluator, if only because it is a much easier task to browse through 100 reconstructions than 2800, and also because of their superficial "high accuracy" as compared to the superficial "sloppiness" of the large macro-etymological dictionaries.

There are, however, two serious problems with this approach. First, claims of "perfect" regularity of correspondences for such a small corpus are usually exaggerated. The very fact that there is only a small handful of data means that most of the correspondences recur only in a few examples, and there is no easy way to statistically verify the significance of these recurrences (as an example, see Starostin 2012 for an analysis of such a situation in the case of Vajda's "Dene-Yeniseian"). Second, if we are speaking of macrofamilies (such as Austronesian with its thousand languages, or Sino-Tibetan with its twenty or more nearly equidistant branches), our main problem is the same as in the case of large etymological dictionaries: a near-limitless choice of comparanda, out of which it is not all that difficult to fashion a "mini-package" of seemingly corresponding forms.

Proposed solution: *Size does not matter*. A long-range genetic relationship hypothesis may operate with as many comparisons as it needs, no more and no less — but only provided there is some sort of objective methodology that helps *range* these comparisons and measure their degree of correspondence to historical and typological expectations. "Hugeness" or "compactness" of the etymological corpus should not be regarded as self-sufficient values — hence, one should refrain from being driven by the desire to either "create as many etymologies as possible" or "purge as many dubious etymologies as possible".

What really *does* matter is our ability to arrange the evidence in a sort of pyramid, where the strongest comparisons (phonetically, semantically, distributionally) should be clearly positioned at the top, and then propped up by as much supportive evidence as is necessary at the bottom. Such an arrangement, among other things, would facilitate external evaluation of hypotheses — one starts by scrutinizing the top part, then selectively examines the rest of the evidence whenever one feels that the "top" evidence is in need of additional supportive data. This is a principle that, surprisingly enough, has not been so far implemented in any of the large etymological corpora that I am familiar with; the closest analogy would probably be M. L. Bender's treatment of comparative Nilo-Saharan evidence (e. g. his system of "excellent", "good", and "fair" isoglosses introduced in Bender 1997, unfortunately, on rather flimsy and poorly stated criteria), but this is more an analogy of style than of substance.

Problem 2: Grammar or lexicon?

Having ruled out extra-linguistic and typological data as evidence for genetic relationship, J. Greenberg (1966) correctly emphasized the importance of grammatical and lexical isomorphisms between compared languages as the main (in fact, only) valid evidence in this particular sphere. On relatively shallow levels of comparison, as is well known to all historical linguists, grammar and lexicon usually go hand in hand inasmuch as the required isomorphisms are concerned. On deep chronological levels, however, this link is just as frequently broken, giving rise to the problem of what is *more* important for demonstrating genetic relationship — grammar (more precisely, paradigmatic morphology) or lexicon (more precisely, *basic* lexicon in a Swadesh-like understanding)? From both a theoretical and a pragmatic standpoint, the seductive power of morphology often wins out, because:

(a) it is assumed, based on selective evidence — never really proven — that morphology is much more resistant to borrowing than basic lexicon, and therefore constitutes a more reliable source for comparison at large time depths (see, e. g., Vovin 2002);

(b) it is relatively easy to prepare and present comparative morphological evidence, just because the amount of such evidence is, by the very definition of morphology, smaller and easier to assimilate than the corresponding amount of lexical evidence;

(c) an additional important point is that, as a rule, comparative morphological evidence is of much greater interest to specialists in typology and various synchronic subdisciplines of linguistics than comparative lexicology — this is yet another stimulus to produce works on comparative morphology, because they do not seem quite as "dull" for the general audience.

At the same time, however:

— argument (a), even if it *is* true, disregards the fact that, while morphology may be more resistant to externally driven change than basic lexicon, it is almost definitely less resistant to internally driven change, where, in addition to general processes of morpheme replacement, we also have to remember the factor of "phonetic erosion" (loss of morphological elements on word borders due to reduction, samdhi, etc.), far more detrimental towards morphology than towards lexicon. For example, as a rule, modern Indo-European languages from different branches tend to preserve far more common morphemes in their basic lexicon than in their morphological systems (no less than 25–30 items on the Swadesh list are still the same, whereas finding 25–30 common Proto-Indo-European grammatical morphemes between, say, a present day Germanic and a present day Iranian language would be quite a challenge);

— argument (b), if taken at face value, may easily lead the researcher into the temptation of thinking along the lines of "I really believe in this genetic relationship, but it is also necessary to convince the public at large, so the best thing to do is to produce some impressively looking morphological evidence, whatever the cost". Analysis of both Laurent Sagart's "carefully assembled" morphological evidence for Sino-Austronesian and Edward Vajda's morphological evidence for Dene-Yeniseian[7] uncovers quite serious subjective distortions of the semantic and distributional properties of the analyzed morphemes, which could have easily been avoided if the researchers in question did not strive so hard to recover the morphological systems (rather than isolated morphemes) of the hypothetical protolanguages.

Proposed solution: *Morphology need not matter*. (Note that this is quite different from "*morphology does not matter*", which would be completely wrong). The presence of strong, cohesive isoglosses that suggest a common paradigmatic morphology for the ancestor of the compared languages is a very strong argument for relationship (on "macro-comparative" levels, chunks of paradigmaticity may perhaps be glimpsed in Niger-Congo and Afro-Asiatic languages, but even in those macro-families the importance of morphology tends to be exaggerated, mainly through "linguistic hearsay"); however, the *lack* of such evidence essentially means nothing (cf. the case of Russian and English, where, other than a tiny bunch of archaic participial suffixes, transparent morphological isoglosses no longer exist).

The reverse, however, is not true: *basic lexicon always matters* — I am willing to risk stating that there is not even a *single* case of commonly accepted genetic relationship, no matter how close or distant, where it would not be possible to demonstrate it on the basis of basic lexicon alone; and in quite a few cases (e. g. the isolating languages of Southeast Asia), it is *only* possible to demonstrate it on the basis of basic lexicon. Consequently, in generating and testing hypotheses of long-range relationship, it is recommendable to always begin with the basic lexicon, which is precisely the ideology behind "The Global Lexicostatistical Database", the Moscow School's latest collective project on language classification. Any long-range comparison that emphasizes morphological reconstruction while downplaying basic lexicon will always look suspicious, since any such scenario would simply violate observed diachronic typology.

Problem 3: Internal solutions or external evidence?

This issue is a rich source for endless debates that usually take place between "narrow" specialists in particular language groups or families and "broader" specialists in macro-comparative studies. The model of the debate is always the same: a Vasconist / Japanologist / Sinologist / Indo-Europeanist / etc. criticizes select long-range etymologies ("Dene-Caucasian", "Altaic", "Nostratic", etc.), concluding that the Basque / Japanese / Chinese / Indo-European / etc. part of the etymology is explainable as secondary on internal grounds, implicitly assuming or explicitly stating that internal etymologization should always take precedence over far-flung attempts at external explanation. Examples are too numerous to list, but see Dybo & Starostin 2007 for a detailed treatment of several such disputes on Altaic etymologies.

[7] For a serious critical analysis of Sagart's "Sino-Austronesian" morphology, see Blust 1995; my own critical remarks on Vajda's "Dene-Yeniseian" morphology have been published as Starostin 2012.

Of all the listed problems, this one is probably the hardest to deal with, because in each individual case there is usually some risk of taking it to almost personal levels — from the point of view of the macro-comparative linguist, the narrow specialist is seen as a retrograde conservative, defending his/her subjective views as a dogma simply because they are supported by "expert authority"; from the point of view of the narrow specialist, it is the macro-comparativist that comes across as an ignorant amateur striving for sensationalism, not interested in real hard work or scientific honesty. (See Manaster-Ramer 1993 for some penetrating insights into the nature of such debates between Nostraticists and their critics).

Proposed solution: This "conflict of interests" may be overcome only when the issue is approached from a thoroughly unbiased position. In particular, flashy, easily memorizable, but overstated phrases like "internal etymologization should always take precedence over external etymologization" should probably be considered just as harmful to doing good science as their opposites.

What is *really* necessary in such cases is an elaborate standard against which the alternatives could be weighed objectively — for instance, a general database of the various types of language change, against the data of which (including statistical data) it would be possible to test the conflicting solutions. One of the most important sections in such a database should be dedicated to semantic change, of which we still know relatively little (e. g. not nearly enough to answer the question, "how likely is it that the noun 'stone' would be derived from the verb 'to roll', anywhere in the world or in one particular linguogeographical area?"). A database of phonetic change might also be important, but only if it is sufficiently detailed — it is very easy to make an inventory for trivial and widespread types of phonetic change, but the toughest problems of etymologization generally tend to concern non-trivial paths of change for typologically rare segmental and suprasegmental units.

In other words, it is necessary, first and foremost, to recognize that most of the debates over internal vs. external etymologization do not so much reveal the personal flaws and biases of the participants (although these things happen, too) as they highlight the weak spots of general comparative methodology; inasmuch as they stimulate us to think of the possible ways to improve the method, they are quite useful, but the important thing is to not let oneself get carried away by ideological, let alone personal, motives.

Problem 4: Formal objectivity or subjective judgement?

This old problem of etymology (the value and importance of subjective judgement and intuition in linguistic reconstruction) has had a major revival in recent times; today, it does not so much pit "lumpers" against "splitters" as it tends to oppose "traditional comparative linguists" and a new generation of scientists — many of them with backgrounds in "hard science" (biology, physics, general computational studies, theory of information etc.) or anthropology and sociology rather than linguistics. What puts them all together is the employment of formal probabilistic methods, usually based on Bayesian principles, to hunt for automatically generated optimal scenarios of language classification (Gray & Atkinson 2003), models of language evolution (Pagel et al. 2007), and, most recently, even protolanguage reconstruction (Bouchard-Côté et al. 2013). Although, in a way, these studies may be said to continue the computational tradition that had already been introduced to historical linguistics by glottochronologists (Swadesh, Lees, etc.) in the 1950s, this is now done on a much more complex scale, with the added potential of high performance computers assisting researchers in selecting the most parsimonious historical scenario.

Traditionally oriented historical linguists seem rather slow to embrace these new methods — not only because many of them have trouble assimilating the mathematical apparatus, but also because even some of those who do not have any such trouble still feel skeptical about whether machine-based methods are powerful enough to achieve results which the human mind cannot achieve on its own (after all, linguistic classification and reconstruction could hardly be reduced to a small set of computational issues). The immediate advantages of formal mathematical methods may seem obvious to the naked eye, but in reality these advantages turn out to be overrated, for the following reasons:

(a) formal methods may give the impression of filtering out the subjective factor — one assumes that they generally produce a mix of correct and erroneous results, and accordingly tries to precisely define the margins of error, something that is rarely, if ever, done in manual etymological research. However, even in the strictest procedures of this kind subjectivity is never really ruled out completely. For instance, any automatic analysis of lists of words/morphemes automatically depends on how accurately these lists have been collected and compiled. Any automatic analysis that tries to build a genealogical tree or a network based on pre-established etymological cognations depends on the degree of subjectivity already present in these arguments. And, finally, any automatic analysis that tackles the data head on, intentionally ignoring all the work that had previously been carried out in the "dark ages" of pre-computational linguistic science, is useless if it does not take into consideration the phonetic, morphological, and lexical specifics of the analyzed languages — something that is more easily said than done;

(b) despite the steady flow of works describing automated procedures, most of them cover relatively "safe" territory — as a rule, the data come from such families as Indo-European or Austronesian, more rarely, Bantu, Semitic, or Turkic. What all or most of these taxa share in common is the following: (a) they have already been well studied by comparative linguists over a research period of 100–200 years; (b) most of them, with the notable exclusion of many modern Indo-European languages, have a relatively simple story of phonetic change, making them ideal "polygons" for testing out simplistic algorithms.

Consequently, we have yet to see actual situations where formalized automatic methods help achieve a real breakthrough in some issue that has proven too tough for "ordinary" comparative linguistics — including, of course, macro-comparative hypotheses as well. Not only that, but even in the respective areas of the listed families application of computational methods has so far been unable to successfully resolve any of the remaining complicated issues: for instance, the most recent publication on the homeland of Proto-Indo-European (Bouckaert et al. 2012) has largely failed to convince Indo-Europeanists of the adequacy of its methodology or the correctness of its results, and the recent publication on the automatic reconstruction of Proto-Austronesian (Bouchard-Côté et al. 2013), has not served to elucidate any of the remaining ambiguities about the phonological system of the protolanguage. Keeping in mind the rather humble efficiency of all these efforts (many of which, I might add, have received the kind of publicity that is quite disproportionate relative to the achieved results), one would, indeed, be seriously tempted to doubt that such methods could be successfully applied to far more complex problems, such as the demonstration and clarification of genetic links between, e. g., North Caucasian languages, let alone Nostratic, Sino-Caucasian, or Amerind.

Proposed solution: Just like in every other situation described above, a compromise is necessary here. Without denying the usefulness and added potential of computational algorithms borrowed from other branches of natural and social sciences and adapted to the needs of historical linguistics, I would suggest that a truly reasonable approach towards any hypothesis of linguistic “macro-relationship” should always strive to combine automated procedures with manually performed research of the traditional type.

Thus, for instance, the ideology that is currently employed in the “Global Lexicostatistical Database” project, which aims at an improved classification of the world’s languages on different chronological levels based on their basic lexicon, recommends taking the following steps:

(a) accurate manual compilation of 100-wordlists for each language group, guided by a formal standard defined in Starostin 2010 and Kassian et al. 2010; (b) generation of two types of cognation indexes — manual (based on phonetic correspondences where they are available and on phonetic similarity where they are not) and automatic, based on the method of consonant class comparison, with different lexicostatistical trees built for both; (c) comparison of results, with well-argumented conclusions on which classification should be regarded as being closer to the truth, and why; (d) reconstruction of the proto-wordlist for the language group; (e) repetition of procedures (a)-(d) for other language groups of the region; (f) comparison of the reconstructed proto-wordlists, and so on (the whole procedure is explained in much detail in Starostin 2013).

The most important role of automatic analysis in this procedure is that in those cases where one ends up with multiple discrepancies between the automatically generated and the manually generated trees, such discrepancies serve as a strong stimulus to look at the data with increased accuracy, so as to understand their nature. In particular, manual analysis may help correct the errors that come from applying the “universal” algorithm of comparison to typologically peculiar situations that were not taken into account at the stage of parameter selection and adjustment. At the same time, automatic analysis may point out some of the subjective weaknesses of “human” etymologization, e. g. reveal cases of fake phonetic irregularity, unjustly postulated by the researcher on the basis of one or two tempting similarities. That way, both approaches complement each other, and regularly switching back and forth between them will unquestionably yield more insights than relying one hundred percent upon only one.

In conclusion, here is a brief list of tasks that I would define as “primary” for the researcher who truly wants to achieve notable progress in the field of linguistic “macro-comparison”:

(1) *Work first and foremost with evidence that may be quantified and statistically assessed*: above all, this means Swadesh-type lexicostatistical lists, but helpful statistical evaluation may also be performed on etymological databases, provided they are complete and organized accurately enough to be easily subjected to various types of automatic analysis.

(2) *Build up reference corpora of typological evidence*: as we move back into the past and multiply our choice options, an important compensatory mechanism for narrowing them back down is correlation with historical-typological databases of the various types of phonetic and semantic change, as well as typological databases that collect and systematize various types of linguistic interaction, primarily loanwords. The construction of a single, unified database of this sort is a major task to be achieved in the future, but even small, “local” databases run by particular teams or individuals may be of use.

(3) *Try to develop and apply universal standards and reference frames*, be it for lexicostatistics (with rigidly defined standard wordlists) or etymology (unified criteria for postulating cognacy). This principle need not be understood simplistically (e. g. as “what is good for Indo-European is also good for Nostratic”), since similar standards should only be applied to similar type objects (Nostratic is, at the very least, much chronologically deeper than Indo-European, so it might require a slightly different approach). But, on the other hand, if every hypothesis of macro-relationship is judged according to its own standard, this renders useless

the best criterion for falsifying such hypotheses — testing alternatives according to the same set of rules. As far as my opinion is concerned, no better universal standard than the lexicostatistical method has been offered so far; however, there is no reason why the lexicostatistical test, relying on a small subset of data, could not be accurately improved in the future by gradually expanding that data.

It goes without saying that not all of the listed recommendations have been properly implemented even for "short-range" families, including such well-studied ones as Indo-European. However, the need to implement them for "long-range" and overall problematic families is much higher — as systems of regular phonetic correspondences between hypothetically related (proto-)languages become harder to establish due to decrease of data and increase of the chronological gaps between different reconstructed states, one must somehow learn to compensate for this added trouble.

At the present time, not a single macro-comparative hypothesis that I know of fully satisfies all of the listed requirements, but some are definitely in better shape than others: e. g. Nostratic fares better in terms of statistical tests and general "standardization" than Austric, while Austric, in its turn, is better than Nilo-Saharan or Amerind, etc. etc. One thing can be said for sure: regardless of whether macro-comparative studies are ever capable of becoming the dominant paradigm in historical linguistics, their possibilities are far from being exhausted, and as long as there are still scholars around who are willing to engage in this exciting field and strive towards "raising the bar", there is really no telling what unpredictable surprises the future may bring.

Literature

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

Blust 1995 — Robert BLUST. An austronesianist looks at Sino-Austronesian // *The Ancestry of the Chinese Language*. Ed. by William S-Y. Wang. Journal of Chinese Linguistics: Monograph Series 8, pp. 283–298.

Bomhard 2003 — Allan R. BOMHARD. *Reconstructing Proto-Nostratic: Comparative Phonology, Morphology, and Vocabulary*. Charleston: Signum Desktop Publishing.

Bouchard-Côté et al. 2013 — Alexandre BOUCHARD-CÔTÉ, David HALL, Thomas L. GRIFFITHS, Dan KLEIN. Automated reconstruction of ancient languages using probabilistic models of sound change // *Proceedings of the National Academy of Sciences*, 10.1073.

Bouckaert et al. 2012 — R. BOUCKAERT, P. LEMEY, M. DUNN, S. J. GREENHILL, A. V. ALEKSEYENKO, A. J. DRUMMOND, R. D. GRAY, M. A. SUCHARD, Q. D. ATKINSON. Mapping the Origins and Expansion of the Indo-European Language Family // *Science*, 337 (6097), pp. 957–960.

Campbell 1988 — Lyle CAMPBELL. Review article on *Language in the Americas* by Joseph H. Greenberg // *Language*, v. 64, pp. 591–615.

Campbell & Poser 2008 — Lyle CAMPBELL, William J. POSER. *Language Classification: History and Method*. Cambridge University Press.

Dolgopolsky 2008 — Aharon DOLGOPOLSKY. *Nostratic Dictionary*. McDonald Institute for Archaeological Research, Oxford Publishing Press.

Dybo & Starostin 2007 — A. V. DYBO, G. S. STAROSTIN. In Defense of the Comparative Method, or the End of the Vovin Controversy // *Аспекты компаративистики III*. Moscow: RSUH Publishers, pp. 119–258.

Ehret 2001 — Christopher EHRET. *A Historical-Comparative Reconstruction of Nilo-Saharan*. Köln: Rüdiger Köppe Verlag.

Gell-Mann, Peiros, Starostin 2009 — M. GELL-MANN, I. PEIROS, G. STAROSTIN. Distant Language Relationship: The Current Perspective // Journal of Language Relationship, 1, pp. 13–30.

Gray & Atkinson 2003 — R. D. GRAY, Q. D. ATKINSON. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426, pp. 435–439.

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

Greenberg 1995 — Joseph H. GREENBERG. The Concept of Proof in Genetic Linguistics // *Mother Tongue I*, pp. 207–216.

Hamp 2011 — Eric HAMP. On The First Substantial Trans-Bering Language Comparison // *The Dene-Yeniseian connection*. Ed. by Jim Kari & Ben Potter, Fairbanks, ANLC, pp. 285–298.

Haspelmath & Tadmor 2009 — Martin HASPELMATH, Uri TADMOR. The Loanword Typology Project and the World Loanword Database // *Loanwords in the World's Languages: A Comparative Handbook*. Ed. by M. Haspelmath & U. Tadmor. The Hague: Mouton de Gruyter, pp. 1–34.

Illich-Svitych 1971–1984 — V. M. ILLICH-SVITYCH. Opyt sravnenija nostraticheskix jazykov. Sravnitel'nyj slovar'. [An attempt at a Comparative Dictionary of the Nostratic languages]. Moscow: Nauka. Vol. 1: 1971; vol. 2: 1976; vol. 3: 1984.

Kassian et al. 2010 — A. KASSIAN, G. STAROSTIN, A. DYBO, V. CHERNOV. The Swadesh wordlist: an attempt at semantic specification // *Journal of Language Relationship*, 4, pp. 46–89.

Manaster-Ramer 1993 — Alexis MANASTER-RAMER. On Illich-Svitych's Nostratic theory // *Studies in Language*, 17, pp. 205–249.

Nichols 1992 — Johanna NICHOLS. Linguistic Diversity in Space and Time. University of Chicago Press.

Pagel et al. 2007 — M. PAGEL, Q. D. ATKINSON, A. MEADE. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, 449 (7163), pp. 717–720.

Sagart 2005 — Laurent SAGART. Sino-Tibetan–Austronesian: an updated and improved argument // *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. Ed. by Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas. London: Routledge Curzon, pp. 161–176.

Starostin 2009 — George STAROSTIN. Review of Lyle Campbell & William J. Poser. Language Classification: History and Method, 2008 // *Journal of Language Relationship*, № 2, 2009, pp. 158–174.

Starostin 2010 — George STAROSTIN. Preliminary lexicostatistics as a basis for language classification: a new approach // *Journal of Language Relationship*, № 3, pp. 79–117.

Starostin 2012 — George STAROSTIN. Dene-Yeniseian: A Critical Assessment // *Journal of Language Relationship*, № 8, pp. 117–152.

Starostin 2013 — G. S. STAROSTIN. *Jazyki Afriki: opyt postrojenija leksikostatisticheskoj klassifikacii. Tom I: Metodologija. Kojsanskije jazyki* [Languages of Africa: an attempt at a lexicostatistical classification. Vol. I: Methodology. Khoisan languages]. Moscow: Languages of Slavic Culture, 2013, 510 pp.

Starostin, Dybo, Mudrak 2003 — Sergei STAROSTIN, Anna DYBO, & Oleg MUDRAK. *An Etymological Dictionary of Altaic Languages*. Leiden: Brill.

Vajda 2011 — Edward VAJDA. A Siberian link with Na-Dene languages // *The Dene-Yeniseian connection*. Ed. by Jim Kari & Ben Potter, Fairbanks, ANLC, pp. 33–99.

Vovin 2002 — Alexander VOVIN. Building a 'bum-pa for Sino-Caucasian: a reply to Sergei Starostin's reply // *Journal of Chinese Linguistics*, v. 30, N 1, pp. 154–171.

Г. С. СТАРОСТИН. Макрокомпаративистика в XXI веке: текущее состояние и перспективы развития.

Статья представляет собой попытку эксплицитно суммировать бóльшую часть теоретических и методологических проблем, на сегодняшний день препятствующих существенному прогрессу в области макрокомпаративистики, изучающей вопросы глубокого родства языковых семей мира. К числу таких проблем относятся, в частности: (а) вопрос о количестве, качестве и природе языковых данных, требуемых для установления глубокого родства; (б) вопрос о приоритетах этимологизации данных при разработке макрокомпаративистских гипотез; (в) вопрос о сложных взаимозависимостях между «объективным» (автоматическим) и «субъективным» (ручным) сравнением. Для каждой из перечисленных проблем в общем виде предлагаются частичные решения или методологические рекомендации.

Ключевые слова: Сравнительное языкознание, дальнее родство языков, языковые макросемьи, вычислительные методы в языкознании.