

The Tower of Babel F.A.Q.

1. The Basics

- 1.1. What IS "The Tower of Babel" and what are its intentions?
- 1.2. What's with the name and what does it have to do with the Bible?
- 1.3. Who runs this place and what are your credentials?
- 1.4. What types of information can be published through the ToB website?
- 1.5. Do you accept contributions from "outsiders"?

2. The Technicalities

2.1. Site Structure

- 2.1.1. The layout is a confusing nightmare! Help!
- 2.1.2. What about the structure of the database list?
- 2.1.3. How DO I make use of the search engine?
- 2.1.4. Okay, I got my query result, but I don't understand anything about it.
- 2.1.5. I need help with your fonts and transcriptions.
- 2.1.6. Can I make use of your databases offline?
- 2.1.7. Why isn't the "DOWNLOADS" section complete?

2.2. STARLING

- 2.2.1. What is STARLING and why should I bother about using it?
- 2.2.2. How do I get STARLING to work on my computer?
- 2.2.3. I installed STARLING, but cannot make head or tails of it.
- 2.2.4. Who do I contact about my STARLING problems?

3. The Science

3.1. The basic principles of our work

- 3.1.1. Why can't I find the language X word I'm interested in in your databases?
- 3.1.2. How *do* you arrive at all these starred forms?
- 3.1.3. Is all this work *yours*? Are there any others out there like you?
- 3.1.4. I find it obvious that Armenian (Hebrew, Japanese, Zulu, etc.) was the proto-language of humanity. Why aren't you paying attention to me?
- 3.1.5. What is the difference between micro- and macro-comparison?

3.2. The macro-comparative issues (*coming soon*)

1. The basics

1.1. What IS "The Tower of Babel" and what are its intentions?

The Tower of Babel project primarily deals with human languages. In regard to these, it can be said to pursue two main goals - Inventarization and Explanation. Both are quite obviously intertwined, and, in fact, most of the participants of the project can hardly imagine their activities as dedicated to one of the two sides exclusively. Nevertheless, ToB can be used as a valuable instrument for those who are only interested in well-established, documentally attested linguistic facts just as it can be useful for those who are willing to move beyond the tried and true into the depths of linguistic reconstruction.

The INVENTARIZATION purpose means that ToB can essentially function as a large collection of linguistic - primarily lexicographic - data, collected from various sources (dictionaries, texts, fieldwork, etc.) and arranged according to a certain genetic classification. ToB, per se, is an etymology-oriented site, which means that the comparative dictionaries presented herein are not usually complete dictionaries of a certain language (although hosting such dictionaries as well *is* a remote perspective). But most of the words that *do* have more or less reliable etymologies are indeed represented in our databases, and quite often, these databases turn out to be just about the only source of lexical information on certain languages, or even language families, that are publicly available on the Web. Obviously this should benefit anyone who is interested in the comparative, typological, or simply trivia-oriented aspects of the world's linguistic diversity.

The EXPLANATION purpose means that much, if not most, of the data presented within the project are not simply presented for collection's sake, but are all part of a large etymological corpus, systematically tied in with each other in order to present, prove, or suggest genetic relationship between the languages involved. Ideally, a project like ToB would be consisting of one (if all the languages of the world go back to a single source) or several (if otherwise) "highest level" databases from which the user could easily descend to any other intermediate level, from large macrofamilies to compact "microfamilies". On practice, since we still know next to nothing about the "Proto-World" language (if it ever existed), such a system is impossible to implicate at the present time. However, most of the etymological databases are still linked together in one way or another, and thus present an opportunity for specialists in the respective field(s) and general users alike to assess much of the work that has been done, over the past decades or recently, in the areas of micro- and macro-comparative linguistics.

1.2. What's with the name and what does it have to do with the Bible?

It is obvious that the name of the site, and of the project in general, can bear specific religious connotations. Therefore, it is important to stress, from the very start, that ToB has nothing whatsoever to do with creationism, anti-evolutionism, or religious approaches to science in general, and that the name is evidently to be taken as little more than a handy metaphor.

The one thing that most of the active participants of the project have in common with the authors of the original "Tower of Babel" myth might be called "a favourable attitude towards the conception of monogenesis", i. e. the idea that all existing human languages must have sprung from an original common source. Even so, "favourable attitude" has nothing to do with "scientific proof", and it can be safely stated that, for the moment, macro-comparative linguistics, despite its progress in the last fifty years, does not possess evidence that would be sufficient to prove monogenesis. On the other hand, neither does it deny monogenesis the status of a valid scientific hypothesis.

1.3. Who runs this place and what are your credentials?

As stated in the introduction, the Tower of Babel website was launched in the mid-Nineties by the late Sergei Starostin, at the time Russia's leading specialist in comparative linguistics and head of the so-called 'Moscow school' of said branch of science. Obviously, this means that the majority of active participants belong to the same school; most of them are either colleagues or disciples of Sergei Starostin, although, of course, this does not necessarily mean that all the authors of ToB always agree about everything, be it theory or practice.

The main bulk of etymological research made public on this site has been so far carried out by the following Russian participants:

Dr. Sergei A. Starostin [1953-2005], former Director of the Center of Comparative Studies at the Russian State University for the Humanities (RSUH); specializing in Indo-European, Altaic, Nostratic, North Caucasian, Sino-Tibetan, Yenisseian studies;

Dr. Anna Dybo, senior researcher of the Center of Comparative Studies (RSUH) as well as the Institute of Linguistics of the Russian Academy; specializing in Indo-European, Turkic, Altaic, Nostratic studies;

Dr. Vladimir Dybo, current Director of the Center of Comparative Studies (RSUH); specializing in Slavic, Indo-European, Nostratic studies, also a major specialist in comparative accentology and tonology;

Dr. Alexander Militarev, senior researcher of the Institute of Oriental and Classical Studies (RSUH); specializing in Semitic, Berber, and general Afro-Asiatic studies;

Dr. Oleg Mudrak, senior researcher of the Center of Comparative Studies (RSUH) as well as the Institute of Linguistics of the Russian Academy; specializing in Mongolic, Altaic, Nostratic, Chukchee-Kamchatkan, and general Paleo-Siberian studies;

Dr. Sergei Nikolayev, senior researcher of the Institute of Slavonic Studies of the Russian Academy; specializing in Slavic, general Indo-European, Uralic, North Caucasian, and Na-Dene studies;

Dr. Iliia Peiros, visiting researcher at the Santa Fe Institute; specializing in Sino-Tibetan, Austro-Asiatic, general Austric, and Indo-Pacific studies;

Dr. George Starostin, senior researcher of the Center of Comparative Studies (RSUH) as well as head of the Department of Far Eastern Philology at the same institution; specializing in Dravidian, general Nostratic, Sino-Tibetan (mainly Chinese), Yenisseian, and Khoisan (Bushman-Hottentot) studies;

Dr. Olga Stolbova, senior researcher of the Institute of Oriental Studies of the Russian Academy; specializing in Chadic and general Afro-Asiatic studies.

Our foreign colleagues and contributors include **John Bengtson** (SFI; specializing in Dene-Caucasian studies); **Merritt Ruhlen** (Stanford University; specializing in Amerind studies, typology, and global etymologies); **Dr. William Wang** (University of Hong Kong; specializing in Sino-Tibetan studies and general historical linguistics and typology); and several researchers from Leiden University (Dr. George Van Driem, R. Rutgers, J. Tolsma) who have generously offered their materials for publication on this website. Naturally, plenty of language material comes from linguistic literature officially published over the last 200 or more years of work, and we try to give everyone credit where possible.

1.4. What types of information can be published through the ToB website?

The first and foremost concern of ToB are etymological databases, created in or converted to STARLING format (see below) and working under the requirements of the STARLING database server. This was how it started, and this is how it hopefully always will be. Of course, not all information is fit for database format, which is why, under the "Articles and Books" section, we occasionally publish some of the textual contributions of our authors - which often serve as guiding lights for databases but are also frequently dedicated to specific topics in comparative linguistics.

However, the linguistic aspect of the world's prehistory is by no means the only one that we are concerned with, since the prehistory of language is, of course, closely tied in with multiple other matters, such as material and spiritual culture, mythology, ethnology, archaeology, and genetics. Therefore, in the future we are hoping to expand the limits of ToB by adding comparative databases on all these subjects as well. Right now, the only such 'new' feature on the site is Yuri Beryozkin's extensive comparative database on mythological motives around the globe - a truly unique source, unfortunately, for the moment only available in Russian.

It must also be noted that Sergei Starostin's interests were never limited to exclusively diachronic linguistics. Among other things, he is the author of a highly successful algorithm for analysis of Russian morphology, and this algorithm, besides being fully implemented within STARLING, is also prominently featured on the site itself. A special section also contains a set of databases dedicated to synchronous analysis and machine-handling of the Russian language (including database versions of Zalizniak's grammatical dictionary and Ozhegov's dictionary). In his later years, his concerns extended to syntax and semantics as well; a side project of ToB is devoted to syntactic analysis of the Russian phrase, and we hope that some of the results of this work (currently headed by Anatoli Starostin) will soon be implemented here as well.

1.5. Do you accept contributions from "outsiders"?

This is an important, yet delicate question; we have received quite a large number of letters from people offering to help, but most of these offers sort of faded away - especially since many of them were, in fact, merely pretexts for asking *us* to help *the writers* of the letters on various issues, from simple consultations (which we are always glad to offer, but many of them can be gotten easily through the site without addressing us directly) to questions like 'Can you tell me where I can get a good education in historical linguistics?' (And, no, we can't. Getting a good education depends a lot on the quality of linguistics' professors and courses, and these things tend to vary and change even among the best universities, so there is no simple answer to this question).

Seriously, though, we are certainly glad to welcome any "outside" assistance. There are about 6,000 or so languages in the world, after all, which is a pretty solid number compared to about 6 or so regular contributors to our database section. Here is how you can help:

- by informing us about the various shortcomings of the site (which are obviously many), including misprints, broken links, content errors etc., either through personal contact or by means of the ToB forum (please be aware, though, that it sometimes takes time to correct them, especially since the site is not really moderated twenty four hours a day by anyone);

- by offering to check, correct, and/or supplement the existing databases if you happen to specialize in the given field and feel like lending a hand;

- by creating (or converting into STARLING format) any etymological databases of your own if you happen to specialize in fields not yet embraced on the site;

- by sending us your work on various issues in comparative linguistics to be put up in .PDF format in the 'Articles & Books' section.

The last three points, of course, require more than just a passing interest in historical linguistics. As far as we are concerned, no one is obliged to have attended the best universities in the world and written PhDs in order to earn the right to do comparative linguistics, but one certainly IS obliged to have a good understanding of the classic comparative method and be at least moderately well versed in the current state of linguistic studies. The ToB team consists of professionals, not amateurs, and we expect nothing less from those who would wish to join it. That said, there is plenty of room for all aspects of historical linguistics on the site, and we heartily welcome participation on the part of specialists in particular families no less than we welcome participation on the part of those interested in long-range comparison.

2. The technicalities

2.1. Site Structure

2.1.1. The layout is a confusing nightmare! Help!

Well, perhaps not exactly so, but a site map would still be in order.

From the titular page (<http://newstar.rinet.ru/main.html>) you can begin browsing the site by clicking directly on the nice Bruegel picture. This brings you straight to the main menu, which is available in both English and Russian versions (which can be switched by clicking on the words 'english' and 'russian' at the top of the page).

The menu options on the left are as follows:

Introduction - a brief overview of the site, also including S. Starostin's original version of the introduction (somewhat obsolete today);

FAQ - you're reading this now;

The Tower of Babel - a brief list of the project's participants;

News - information on updates and upgrades;

Languages of the World: Etymological Databases section:

All Databases - brings up a scrollable list of available databases;

Interactive Maps - brings up a list of maps of the world's language families from which you can also access the databases;

Download - opens the database section of the download page.

STARLING section:

What Is It? - a brief description of the program and its possible use;

Download - opens the STARLING section of the download page.

Articles And Books - a list of electronically available text works from ToB participants, classified by (macro)family and available either in .PDF or .HTML format (or both).

Links - links to other useful linguistic resources on the Web.

Technical Advice - what it says, but the information on the page primarily concerns *fonts*, since all the databases are extremely 'transcription-heavy' and getting them to look right on your computer may be a real pain without exact instructions.

The Analytical Catalogue of Mythological Motifs - see section 1.4. Available in the same database format as the etymological files. Unfortunately, at the moment this huge database is only available in Russian. If you would like to contribute to making it available in English, give us a call.

In memoriam: S. Starostin - a small section dedicated to cherishing the memory of ToB's creator. Contains biographic material, friends' and colleagues' reminiscences, photos, and even a few audio files. Most of the section is in Russian.

Forum - a place to contact us.

2.1.2. What about the structure of the database list?

The list is structured as a hierarchy, progressively going from the highest possible level ('global' etymologies) to the lowest levels of small linguistic groups. A typical entry might look something like this:

Tungus etymology

Compiled by Anna Dybo [view](#) | [query](#) | [description](#) 2006-02-14

The legend is as follows:

Tungus etymology: name of the database (in a compact form for aesthetic purposes, the full name should probably read something like 'An Etymological Database of the Tungus-Manchu languages').

Compiled by Anna Dybo: name of the project participant who takes primary credit for the compilation of the database and computerization of the material. NB: "compiler" does not necessarily mean "author of the etymologies". E. g., the authorship of J. Pokorny's Indo-European dictionary obviously belongs to J. Pokorny; "compiled by G. Starostin" merely means that the latter was responsible for transferring the data from book to database. In a lot of cases, though, "compiler" is identical with "author", and this is explained in the database description.

View: transfers you to the alphabetical listing, page by page, of the database.

Query: transfers you to the search page (see 2.1.2).

Description: transfers you to a brief description of the database. Not all of the descriptions are currently online (we're working on it), but an ideal description would include (a) a brief explanation of what the database is about, (b) a rundown on the contents of each particular field in the database, (c) authorship and sources of the database, (d) notes on transcription, (e) phonetic correspondences between the compared languages. Most of the descriptions, unfortunately, still include only some of these points.

2006-02-14: the date when the online database was last updated (*note:* not all the updates are immediately uploaded on the site, so it is advisable to contact the compilers in person if you want to refer to the most recent version of the database).

Additionally, some of the databases (mostly 100-wordlists) also contain the option **Tree** that takes you to the genealogical tree of the corresponding family, constructed automatically by the STARLING software on the basis of lexicostatistical calculations.

2.1.3. How DO I make use of the search engine?

It's really much more simple than it looks. Let us illustrate with a few concrete examples.

a) *I want to know how many words for "snow" there are in Proto-Eskimo.*

From the main menu column on the left, select "all databases", then find "Eskimo etymology" in the list. Click 'query'. In the field 'meaning', enter 'snow' (without quotation marks). Then press the 'Search' button at the bottom. You're all done!

Granted, you also got the Proto-Eskimo word for 'snowshoe' in the process because the string 'snow' was part of that word. If you only want to get results for whole words (and you probably do), check the 'Whole Words' box next to the 'Meaning' field before submitting your search.

If you only want to have meanings that *begin* with 'snow' (i. e. snow (fallen)', but not 'to be covered with snow'), what you need to do is click on the 'Query method' menu next to the 'Meaning' field and select 'Match beginning' before submitting the search.

b) *I want to look up the etymology of the English word 'louse'.*

Select "all databases", go to "Germanic etymology", click 'query'. In the field 'English', enter 'louse' (without quotation marks). That's all there is to it. [Please note, though, that only English words with reliable Proto-Germanic etymologies will be found - as this is an etymological database for Proto-Germanic, not English. Thus, it will be useless to enter words like 'apple' or 'orange' or 'constitution' since they are not of Proto-Germanic origin. The same principle is applicable to most other languages in our databases, with the possible exception of Russian for which a separate etymological dictionary database - Max Vasmer's - is available].

c) *I wanted to look up the word for 'eye' in the Kagoshima dialect of Japanese and got 'no records found'. How come? Such a basic word and it is not in the database?*

What you did is you probably went to "Japanese etymology", selected 'query', found the field 'Kagoshima', entered your desired meaning and submitted the search. **THIS SHOULD NOT BE DONE.** That field in the database only contains the *form* of the Kagoshima words but not their *meanings*. Only enter the meanings in fields which are explicitly called 'Meaning' and nowhere else. So, what you should have done is enter the string 'eye' in the field 'Meaning'. This immediately gives you the Proto-Japanese form for 'eye' - along with all the reflexes in daughter dialects, including Kagoshima.

d) *I want to enter a word containing special symbols but don't know how to do it.*

Alas, it is for now unclear how to implement the mechanism of 'advanced keyboard input'. Currently, there are two solutions. You can type the word in an external editor (Starling, or, in some cases, Microsoft Word) and use the copy-paste mechanism to put it in the query window. Alternately, you can input it in a simplified way (e. g. type *panca* instead of *pāñca* for the Old Indian word for 'five') and select the 'Like substring' option in the query method. This will give you all the words whose basic phonetic shape will look like *pāñca* - which, in this particular case, will only include one word (*pāñca* itself).

2.1.4. Okay, I got my query result, but I don't understand anything about it.

The basic thing to remember about ToB databases is that they represent a tightly bound hierarchic structure. So chances are, whenever you submit a query, you are not getting the full amount of etymological information about your word, but only a part of it.

Let us take further the 'louse' example. When you look up the etymology for English *louse*, what you get in return looks as follows:

Proto-Germanic: *lūs

Meaning: louse

IE etymology: 

Old Norse: lūs, pl. lūss f. `Laus'

Norwegian: lus

Swedish: lus

Danish: lus

Old English: lūs, pl. lūs f. `louse'

English: louse

Middle Dutch: luus f.

Dutch: luis f.

Middle Low German: lūs

Old High German: lūs (9. Jh.)

Middle High German: lūs st. f. 'laus'

German: Laus f.

The word in bold green letters is the name of the parent family (in this case - Germanic), with the proto-reconstruction next to it (*lūs). Below you get all the related forms in Germanic languages, including English (the string you submitted as your query is always highlighted in red letters). However, Proto-Germanic *lūs 'louse' is not without an etymology of its own; it goes back to a Proto-Indo-European form which has parallels in Indo-European languages from other subgroups.

Therefore, if you want to access the higher level of reconstruction, look for the little hand symbol; in this case, it is sitting next to 'IE etymology'. If you click on it, the screen will change and what you get is this:

Proto-IE: *lewə-

Nostratic etymology: 

Meaning: louse

Russ. meaning: НАСЕКОМОЕ (ВОШЬ)

Baltic: [*lū-l-iā f.](#)

Germanic: [*lūs- c.](#)

Celtic: OCorn lowen, NCorn lūan, pl. lūan `Laus'; OCymr leu-esicc `von Läusen zerfressen', NCymr lleuen, pl. llau, Bret laouen `Laus'

References: [WP II 443](#)

This is the corresponding entry in the Indo-European database. What does it have?

- a) Proto-IE: the Proto-Indo-European reconstruction.
- b) Nostratic etymology: see below.
- c) Meaning: meaning (in English).
- d) Russ. meaning: meaning (in Russian).
- e) Baltic: the Proto-Baltic reflexation of the Indo-European root, linked to the Baltic etymology database.
- f) Germanic: the Proto-Germanic reflexation, which we have already examined.
- g) Celtic: select forms in Celtic languages that also go back to the same root. Ideally, this field should link to a Celtic database, but one has not yet been prepared; instead, you just get a selection of forms all in one field.
- h) References: list of printed (or unprinted) sources. In this case, there is only one major source, the etymological dictionary of Walde-Pokorny. References are usually given in an abbreviated way; they can be deciphered either by clicking on them to open the bibliography database, or by consulting the general description of the database by clicking on 'description' in the main database list.

From here, you have the following additional choices:

- see in more details the reflexes of PIE 'louse' in Baltic languages. This can be achieved by clicking on the Baltic proto-form that is in fact a hyperlink in itself. This takes you on a lower level of reconstruction (hierarchically the same as Germanic);
- conversely, you can go higher and check out the hypothetical Nostratic roots of this Proto-Indo-European root. Click on the little hand icon next to 'Nostratic etymology' and you will open a

new window, this time for Proto-Nostratic *IVŋV 'a kind of insect', with possible parallels in Altaic, Uralic, Dravidian, and Eskimo languages;

- finally, you can select the '**List with all references**' option below. This will give you an exhaustive list of *all* the parallels for Indo-European 'louse': both on the lower levels (Germanic, Baltic), the higher level (Nostratic), and the lower levels of sister families (so you will also get entries from the Altaic, Uralic, Dravidian, Eskimo etc. families and their sub-families as well - quite a long list).

The bottomline is you can make a really long and exciting etymological journey that might yield exciting results (such as discovering, for instance, that English *louse* may be ultimately descended from the same source as Chinese *chóng* 'insect!'). Not all the etymologies are of equal probabilistic stature, of course, but different degrees of reliability are unavoidable in any etymological work.

2.1.5. I need help with your fonts and transcriptions.

The first thing to realize is that most of the data simply cannot be displayed well without special fonts; too many languages are involved to make it all easily coverable by a standard Times New Roman or Arial.

If, for some reason, you are vehemently adamant about installing a different font - or if there are some insurmountable technical problems about it - there is an option called 'Recoding' which tries to convert all the special symbols into conventional ASCII combinations. You may try it any time from within a database by clicking on '*Change viewing parameters*', checking the box next to '*Recoding*' and then selecting '*Change*'. This will make everything readable, but rather ugly, so you should probably only resort to this option in extreme cases.

A slightly more difficult, but far more fruitful, way to get out of this is to download the Times New Roman Star font from this link: <http://starling.rinet.ru/timestr.zip>, unzip it (WinRar or any other standard archiver software may be used) and install it on your PC. The font contains all standard Roman characters, Cyrillic characters and special phonetic characters. The fonts for Macintosh (coding table: Western MacRoman) can be downloaded [HERE](#).

2.1.6. Can I make use of your databases offline?

Yes, of course, and in fact you are *encouraged* to do so, especially if you have more than just a passing interest in comparative linguistics and etymology and need the databases for constant rather than occasional references.

Many of the databases are available in download form on the "Downloads" page; you need only to upload the self-extracting .EXE archives to your computer and run them from within Windows to extract them to any specific folder you want. The only catch is that, since they use STARLING fonts and a special STARLING format convention called 'variable length fields', they will not be usable by any other database viewer and editor than STARLING itself. Therefore, in order to make everything work, you will also have to download and install STARLING on your computer (for instructions, see below).

Once everything is installed and running, you can modify the databases according to your own wishes and preferences.

PLEASE NOTE that if you want to make use of the databases for your academic research, references to the specific database's Internet address (<http://starling.rinet.ru>) and author are obligatory. In addition, it is highly recommendable that you inform the author by E-mail, as it is always possible that a newer version of the database, which the author might be willing to share, is already

available (after all, most of them represent work in progress) but has not yet been uploaded to the server for technical reasons.

2.1.7. Why isn't the "DOWNLOADS" section complete?

Meaning, of course, that many of the databases that are well viewable online are not publicly available for offline viewing. The reason for this is that many of the bases are very much in a "state of progress", suffering from expectable problems like raw data, poor documentation of sources, various hard-to-understand notes and abbreviations, etc.; some, moreover, are drastically incomplete. Our position here is simple: their Internet presence is useful in that they give the general viewer an overall perspective of what has been done by the team and what is still in store, but it is not altogether recommendable to use them as a highly reliable source for external research.

Therefore, only databases that we view as representing more or less "final products" - e. g. Sergei Starostin and Sergei Nikolayev's North Caucasian dictionary, the Altaic etymological dictionary, etc. - are available for full download. As time goes by, more will be added to the downloads section; but it should be understood that generally this section will be updated much more rarely than the regular online one.

2.2. STARLING

2.2.1. What is STARLING and why should I bother about using it?

STARLING is a software package originally designed and programmed by the late Sergei Starostin (hence the name) since as early as the late 1980s. Its most important goal is to offer linguists a suitable and adjustable working environment for creating linguistic databases, primarily etymological and etymostatistical ones, and using them to analyze language data for various historical (and not only) purposes. In addition to that, STARLING has many incorporated features - such as text editing, printing, etc. - but its exclusiveness is provided by its rather unique database interface and a series of linguistic analysis functions unavailable elsewhere.

One major reason for downloading and installing STARLING is that otherwise, you will NOT be able to view and make use of our databases offline. For a series of technical reasons - most importantly, the use of variable length fields and non-standard fonts - they will be incompatible with the usual dBASE format. STARLING does, however, offer full convertibility of its databases and texts to such popular external formats as .XML (for databases) and WinWord (for texts), although, of course, none of the linguistic analysis procedures can be implemented after such a conversion.

If you wish to conduct some linguistic research of your own, we strongly encourage you to make use of the possibilities offered by STARLING. Its vast array of search, filtering and indexing functions, as well as the inbuilt apparatus of lexicostatistic analysis for Swadesh wordlists, makes it a very valuable tool. Drawbacks include an interface that is still not as user-friendly as it could be (although we're slowly working on it) and occasional bugs and flaws that are gradually corrected based on user feedback.

2.2.2. How do I get STARLING to work on my computer?

Provided you have an IBM PC with Microsoft Windows (v. 95 or higher) installed, that is not very difficult. All you really have to do is to download the STAR4WIN executable archive from the 'Downloads' section, double-click on it and let it be installed into the folder of your choice (although the default choice, \Starsoft\, is always preferable).

Alternatively, you can also download and install the STAR4DOS package, which is exactly what it is - a version of Starling designed for usage under Microsoft DOS. The basic advantage is, of course, that it will work on older versions of Windows and under DOS, provided you are still using them. The disadvantage is that the DOS version is no longer supported by our programming team and thus lacks a number of features recently implemented in Star4Win - although, of course, any database created and maintained in the DOS version will be fully compatible with the Windows one.

Additional "extension packages" for the Windows version are also available in the 'Downloads' section. These include STARCJK.EXE, a special self-extracting archive containing additional CJK fonts for handling Chinese characters (please note that you will have to reboot your PC after the installation); DICTS.EXE, the archive containing Zalizniak's Russian dictionary and Mueller's English-Russian dictionary for those who will need STARLING for its morphoanalytical capacities; and MULTILANG.EXE, an archive of a European multi-language dictionary which can be used for getting rough translational equivalents of a lot of stuff.

Finally, a STARLING version for Apple McIntosh is currently in the final stages of preparation and we hope to be able to offer it for download in the nearest future.

2.2.3. I installed STARLING, but cannot make head or tails of it.

Yes, STARLING is definitely not the most user-friendly program ever made - it has always valued substance over form, primarily because for the first ten years or so of its existence, the only person working on it was Sergei Starostin himself. We are working on making it more so, but it sure takes a long time. In the meantime, there is a HELP section included in the program which takes care of some of the trickier aspects of the program, although it is far from covering all of its abilities.

That said, if all you really want is to browse through downloaded databases, it is quite simple - all you have to do is choose "File > Use for BROWSE" or, for an alternate view, "File > Use for EDIT" in the menu and select the location of the database.

2.2.4. Who do I contact about my STARLING problems?

For specific help, you can contact either George Starostin (gstarst@rinet.ru) or Phil Krylov, the lead programmer for STARLING (phil@newstar.rinet.ru), or, alternatively, leave your question on the ToB Forum for all to see.

3. The Science

3.1. General questions

3.1.1. Why can't I find the language X word I'm interested in in your databases?

The ToB databases are, in essence, a set of *etymological dictionaries of language families* - lists of roots and stems of reconstructed languages (such as Proto-Indo-European, Proto-Altaic etc.) accompanied by their reflexations in daughter languages, including modern ones. With a few exceptions (the most notable being our computerized version of Max Vasmer's Russian etymological dictionary), they are not structured as *etymological dictionaries of languages*.

What this means is that if you are looking for the origins of the English word 'origin', for instance, you will not discover it by going to the Germanic database, typing that word within the field 'English' and then asking the machine to search for it — because that particular English word is of Latin, not Germanic, origin, and is not present in Proto-Germanic stock.

Likewise, in most cases you will probably not find etymologies for derivatives; looking for the etymology of Eng. 'louse' will yield positive results, but doing the same for Eng. 'lousy' or 'delouse' most certainly will not.

Of course, in many cases the explanation is less pleasant: namely, many of the databases are incomplete (some, particularly in the Afroasiatic part of the site, may incorporate less than a hundred different roots). This situation, however, will eventually be remedied.

3.1.2. How *do* you arrive at all these starred forms?

By following the standard methodology of comparative historical linguistics, described and elaborated by several generations of Indo-Europeanists over the XIXth century. The basics of this methodology can be looked up in multiple locations (starting from Wikipedia and ending with any general textbook on historical linguistics, regardless of the author).

Several caveats should be mentioned, though:

a) While the actual linguistic data on the site never really change (with the understandable exception of correcting misprints or, sometimes, switching from a less reliable source to a more reliable one), reconstructions *do* change, and sometimes change frequently, particularly when the database in question is in its initial stages of preparation. Thus it is always useful to check for newer versions of the databases.

b) Not all the 'reconstructions' are of the same quality. Some of the starred forms aren't really reconstructions at all, but rather representations of some common 'invariant' of a number of different transcriptions (e. g. the Hadza and Sandawe databases, where there is only one language involved and nothing really to reconstruct). Some are based on forms attested in only one or two languages of a given group, which obviously makes them less reliable - although the presence of external parallels on a higher genetic level sometimes boosts their credibility. Finally, some of the starred forms (this relates particularly to the 'global etymology' database) are not based on regular correspondences but rather on phonetic similarity, which makes them merely an approximate model. In the light of this absence of a unique standard, it is always necessary to check the general description of the database (where available) before making use of it.

3.1.3. Is all this work *yours*? Are there any others out there like you?

Certainly not all of this work is ours - in that much of it is based on the research, both previous and current, of qualified specialists in select fields of historical linguistics. Thus, the Indo-European database makes heavy use of the etymological dictionary of A. Walde & J. Pokorny (which, in turn, was based on decades of hard work by dozens of other scholars); the Uralic database rests upon K. Redei's comparative dictionary; the Nostratic database builds upon the legacy of V. M. Illich-Svitych, etc. etc. Credit is regularly given to all these scholars where necessary, in the form of general additions to the database description or particular references in the appropriate fields (many of them accompanied with a hyperlink to the general bibliographical database, although this is still very much in progress).

As for anything on the *Web* that resembles our work, so far we are not aware of any global resource that closely approaches "The Tower of Babel" either in scope or in methods and goals. Two of the most prominent sites that feature a global approach towards the world's languages are the [Ethnologue](#) (run by the Summer Institute of Linguistics) and [The Rosetta Project](#). However, neither of them hosts any historical databases; the former is dedicated to building up a classified inventory of the world's languages, while the latter simply functions as a select data archive.

Occasional sites on historical databases of select language families can be sometimes met on the Web (some are listed on the Links page), but none are as expansive as ToB.

3.1.4. I find it obvious that Armenian (Hebrew, Japanese, Zulu, etc.) was the proto-language of humanity. Why aren't you paying attention to me?

For a good reason. Both micro- and macro-comparison in historical linguistics (see 3.1.5) should be strictly distinguished from *pseudo-comparison*, cases of which are, unfortunately, not unfrequent, especially among people who, either through lack of sufficient knowledge or certain other personal reasons, believe that intuition alone is able to explain the similarities between the world's languages.

The Tower of Babel is a strictly scientific project, and, as such, has zero tolerance for pseudo-comparison. It would therefore probably be useful to offer, very briefly, a few general guidelines on how to separate real historical linguistics from pseudo-linguistic wizardry, especially for those readers and users who do not have a solid background in this field, but may nevertheless show general interest in it:

1) At the heart of comparative linguistics lies the basic principle of *regular phonetic correspondences* between the compared languages which allow us to set up a number of strict phonological rules deriving daughter languages from their ancestor. Pseudo-comparison, on the other hand, almost always deals with the idea of *phonetic resemblance*, so that any word in language A which is phonetically similar to a certain word in language B are deemed comparable.

2) Comparative linguistics makes serious use of linguistic typology, taking into account the types of historically attested phonetic and semantic developments as well as typologically frequent or universal language features, so as to treat the reconstructed proto-language as a realistic system rather than an artificial construct. Any such proto-languages, reconstructible on the basis of historically attested idioms, do not significantly differ from their descendants. On the contrary, pseudo-comparison is often preoccupied with the questions of glottogenesis, deriving all the words of the world's languages from a few "primal" elements and "reconstructing" the Mother Tongue as merely a combination of such. This has nothing to do with true comparative studies.

3) Comparative linguistics tends to take as much data as possible into consideration; ideally, the system of intermediate and final reconstructions should be able to account for all the language material that is documentally witnessed. The procedure of comparison is regularly conducted from lower to higher levels. E. g., it is methodologically incorrect to compare Modern English to Modern Chinese; the comparison should stretch from English to Proto-Germanic (accounting for all the different stages of English as well as related Germanic languages), from Proto-Germanic to Proto-Indo-European (taking into account the data from other branches of this family), from Proto-Indo-European to Proto-Nostratic, and from there to Proto-Borean or Proto-Eurasian. For Chinese, in its turn, it is vital to reconstruct its Old Chinese ancestor (based on dialectal as well as hieroglyphic data), compare it with other Sino-Tibetan languages (Tibetan, Burmese, etc.), then proceed to Sino-Caucasian, and from Sino-Caucasian to Borean again. Any other comparison between English and Chinese is more or less meaningless and makes it impossible to distinguish between true old cognates and numerous chance resemblances.

3.1.5. What is the difference between micro- and macro-comparison?

The Tower of Babel is a project launched and run by people who are commonly designated as "lumpers", i. e. historical linguists who are highly interested in long-range comparison - establishing genetic relationship between distantly related languages and language families and reconstructing the corresponding protolanguages.

That having been said, we must necessarily insist that there is no intrinsic methodological difference whatsoever between "micro-comparison" (Indo-European linguistics, Uralic linguistics, Semitic linguistics, etc.) and "macro-comparison" (Nostratic, Sino-Caucasian, etc. studies). The only thing that is different is that in the case of micro-comparison, it is either modern day languages or extinct, but documentally attested languages that constitute its primary object; in the case of macro-comparison, this position is commonly occupied by protolanguages reconstructed through micro-comparison (i. e. Proto-Indo-European is reconstructed on the basis of Old Indian, Old Greek, Latin, ancient Germanic and Slavic languages, etc., etc., whereas Proto-Nostratic is reconstructed on the basis of Proto-Indo-European, Proto-Uralic, Proto-Altaic, Proto-Kartvelian, and Proto-Dravidian).

However, since it is generally assumed that the typological features of the reconstructed proto-languages were not at all different from the nature of modern day languages - at least, there is nothing in the obtained "micro-level" reconstructions that would convince us otherwise - it is clear that obtaining deeper level results from comparing intermediate proto-languages should be conducted according to the same procedure that is used in micro-comparison: that is, the classic comparative method which has not lost any of its validity since its elaboration in the XIXth century.