

# Are you my mother . . . Tongue?

by *Daniel Rockmore*



CONSTRUCTION OF THE TOWER OF BABEL,  
HENDRICK VAN CLEVE III, CA 1525-1589  
PHOTO: SNARK/ ART RESOURCE, NY

**T**he story of the Tower of Babel is the creationist's version of the origin of language diversity: Man, in one of his many hubristic moments decides to build a tower to Heaven. God, realizing that communication is the key to completing any massive public works project, foils the plan by replacing the single common language of the workers with many different languages, thereby making impossible their cooperation, not to mention the scheduling of car pools and the organization of a softball team.

Is there a true "mother tongue" to which all existent modern languages can trace their origin? More precisely, what is the genealogical tree or phylogeny of language? It is these sorts of questions, the ones that look to tell a story of a branching journey of the development of languages, that is the goal of SFI's Evolution of Human Languages (EHL) Project, funded by the John D. and Catherine T. MacArthur Foundation, and spearheaded by SFI Distinguished Fellow Murray Gell-Mann and Russian Academy of Sciences

Member (and frequent SFI visitor) Sergei Starostin. The third leader of the project is Dr. Merritt Ruhlen from Stanford University, author of the monographs: "Guide to the World's Languages" and "The Origin of Languages."

The EHL project falls squarely within the discipline of comparative linguistics. The last 200 years or so of the subject have been devoted to the clarification of the most elementary stages of linguistic organization, an effort that has resulted in a partitioning of the roughly six thousand attested languages into several hundred more fundamental "language families," each of which implies the existence of a single language ancestor for its family members.

The standard methodology used to show relatedness involves the identification of a set of phonetic similarities between the words in the respective basic vocabularies (e.g., words for body parts, numerals, natural phenomena, etc.). This is the sort of comparison that supports the existence of a common Germanic language able to account for the English







same probability of replacement to each word in the basic vocabulary. This model is quite naïve and Gell-Mann is leading an effort directed toward tuning the model using more realistic estimates of replacement probabilities of individual words.

Current techniques appear to reliably reconstruct the “protolanguages” in use six to seven thousand years ago. In addition, there is striking evidence for the existence of about ten “superfamilies” responsible for all languages in use today. The analysis reveals some interesting family relations; for example, it indicates that Northeast Asian languages such as Korean and Japanese are closer to European languages than are Southeast Asian languages (e.g., Chinese).

The theoretical (i.e., model building) component of the EHL program is paired with (if not made possible by) a huge empirical component. A part of this component stems from the recently completed *Etymological Dictionary of Altaic Languages*, which gives a comparative study of the Altaic languages, for which Starostin is co-author. See facing page.

The print component of the project is important, but the process of comparison and modeling is primarily focused on the development and management of a growing collection of online language databases. At the head of the database effort is Starostin, whose software package “STARLING” is designed specifically for linguistic database management (see <http://starling.rinet.ru>). The number of online language databases is increasing steadily. Of

great current interest is the effort to digitize all the languages of New Guinea, an effort that will go a long way toward the reconstruction of the Indo-Pacific protolanguage.

The installation at SFI of the entire STARLING project (software, webserver, etc.) is one of the major directions of current work in the EHL project. This, in concert with the mathematical modeling effort, defines the EHL project as another cornerstone in SFI’s work at the scientific frontier. Our generation is bearing witness to a long-overdue mathematicization of the life and social sciences, a modern updating of the Tower of Babel tale in which, through the ever-broadening mediation by the universal language of number, scientific knowledge is growing via a renewed unification across disciplines. SFI’s work to find the Mother Tongue is yet another instance of the progress propelled by the rewriting of sciences in the Mother Tongue of mathematics.

*Daniel Rockmore is professor of mathematics and computer science at Dartmouth College and a member of the SFI External Faculty.*

ILLUSTRATION: SNARK/ART RESOURCE, NY