

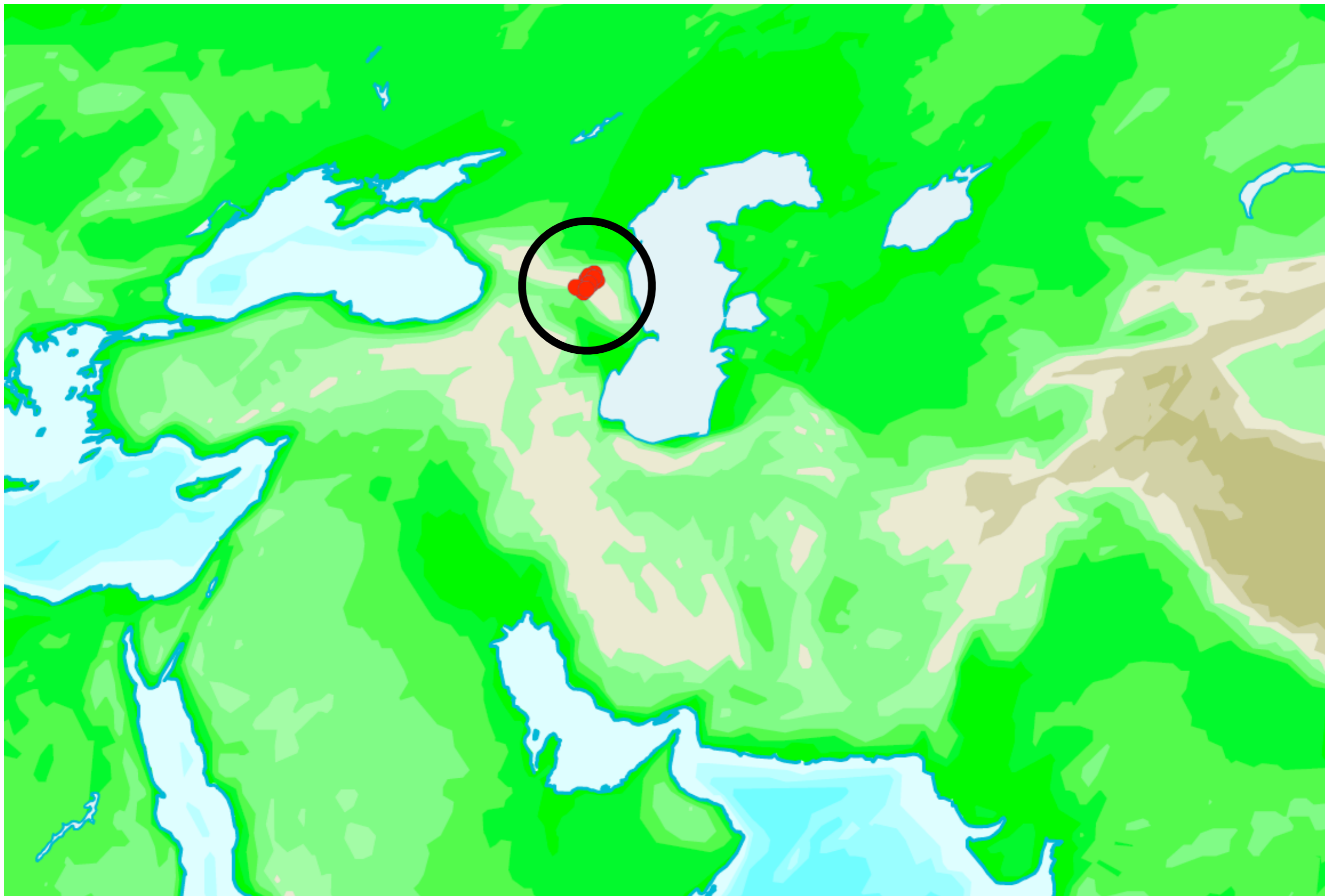
Exemplar-Based Language Change

Michael Cysouw

Research Unit “Quantitative Language Comparison”

Background

- ERC ‘Starting Grant’:
Quantitative modeling of historical-comparative linguistics:
Unraveling the phylogeny of native South American languages
- Data-mining all those sources out there
- Digitization of relevant printed dictionaries



Tsezic

West Tsezic

East Tsezic

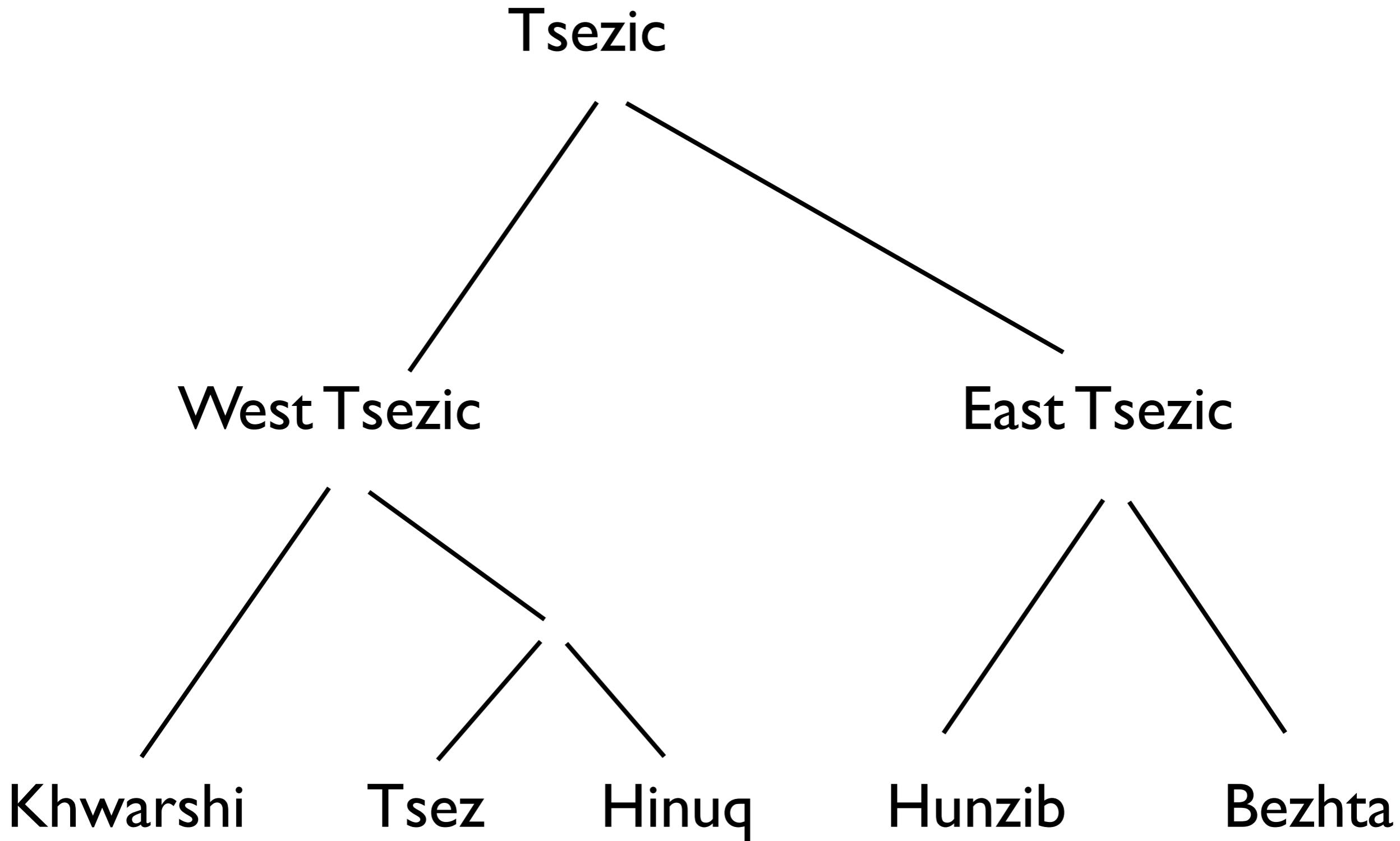
Khwarshi

Tsez

Hinuq

Hunzib

Bezhta





[IDS Main Page](#)
[Simple Browsing](#)
[Advanced Browsing](#)
[Download Data](#)
[Technical Background](#)

The Intercontinental Dictionary Series

Founding Editor:

Mary Ritchie Key (University of California, Irvine)

General Editor:

Bernard Comrie (Max Planck Institute for Evolutionary Anthropology, Leipzig)

Purpose: The purpose of the IDS is to establish a database where lexical material across the continents is organized in such a way that comparisons can be made. Historical studies, comparative, and theoretical linguistic research can be based on this documentation. This is a long-term cooperative project that will go on for the next generation or so and will involve linguists all over the world. It is aimed towards international understanding and cooperation. This is a pioneering effort that will have global impact. The purpose also contributes to preserving information on the little-known and "non-prestigious" languages of the world, many of which are becoming extinct.

Rationale: Information on languages of the world is scattered over all the continents and islands and published in dozens of languages and scripts. There is need of a database where one can find comparable material to formulate hypotheses and test and validate those theories. For example, theories on intercontinental connections have been proposed on the basis of the distribution of 'sweet potato' and yet there is no single source, where words with this meaning can be found in many languages. Good quantitative and statistical studies are almost impossible to do now in non-Western languages. The IDS will provide a quantitative base for a scientific approach to language analysis and comparisons. The IDS will provide the research tools necessary for expanding studies such as phonological theory, word formation, language change, lexical distribution, symbolism and onomatopoeia, classification, and other ideas that have to do with history of people and migrations. The IDS will serve not only as a synonym dictionary but as an index to meaning and to cultures of various people around the earth.

Plan of Series: The IDS series may appear as 1) a volume with 25 or more languages recorded; 2) a fascicle with 5 to 10 languages recorded; 3) in single WordLists, which are archived until enough are gathered to make up a fascicle or volume. A list of fascicles and volumes in progress is available from the general editor.

Procedure: The IDS is developed in cooperation and complementation with other research projects. Throughout the world there are linguistic activities from establishing of databases in universities and think-tanks to publishing grammar series and literacy materials, to individual projects such as the Tibetan dictionary project. Many projects seek to make linguistic data accessible in a format that will allow generalizations to be made. The computer now gives us the potential for tying together linguistic databases. The IDS editors will continue to monitor linguistic activity around the world, both for choosing the languages for forthcoming compilations and for collaboration with other research teams.

<http://lingweb.eva.mpg.de/ids/>

Exemplar-based research

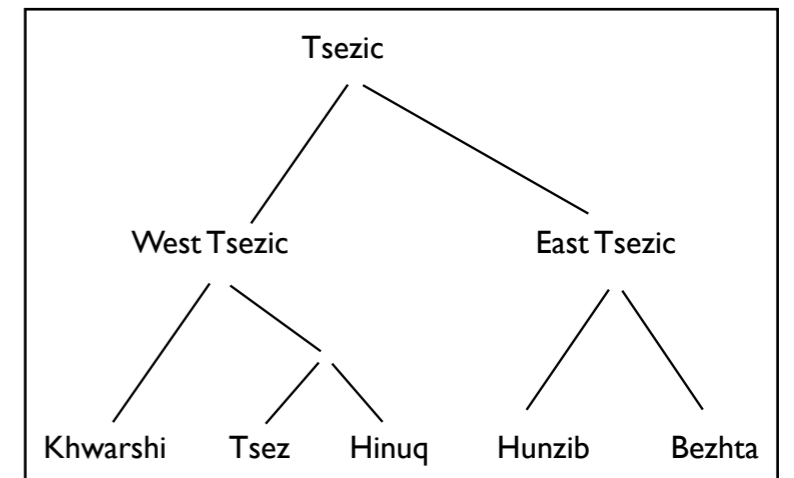
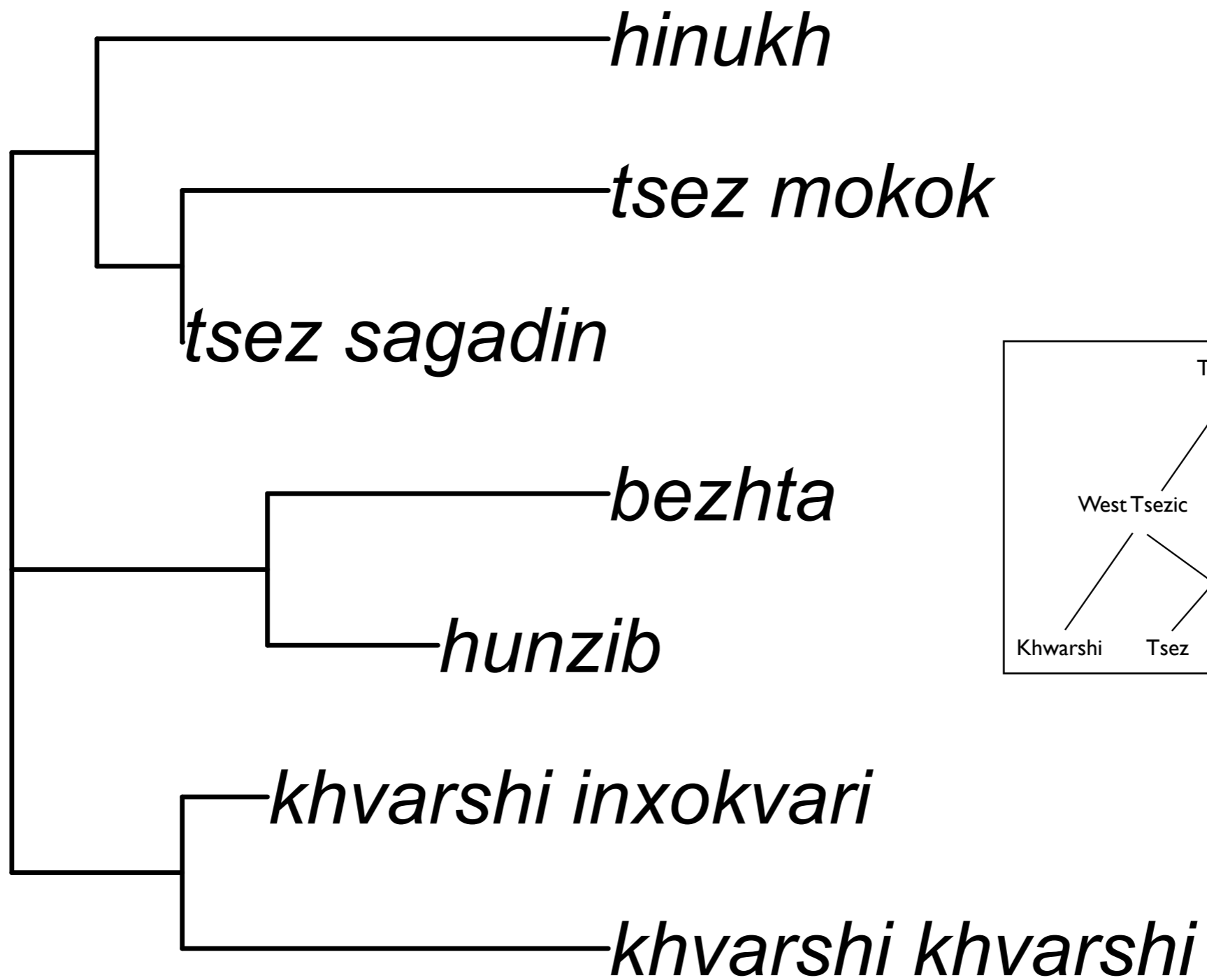
- Token-based, not type-based
- Use distributions of measurements, not idealized 'averages' (point estimates)
- Compare sets of utterances, not abstract representations

Exemplar-based research

- phoneme ~
set of concrete utterances in context
- meaning (sememe) ~
set of concrete usages in context

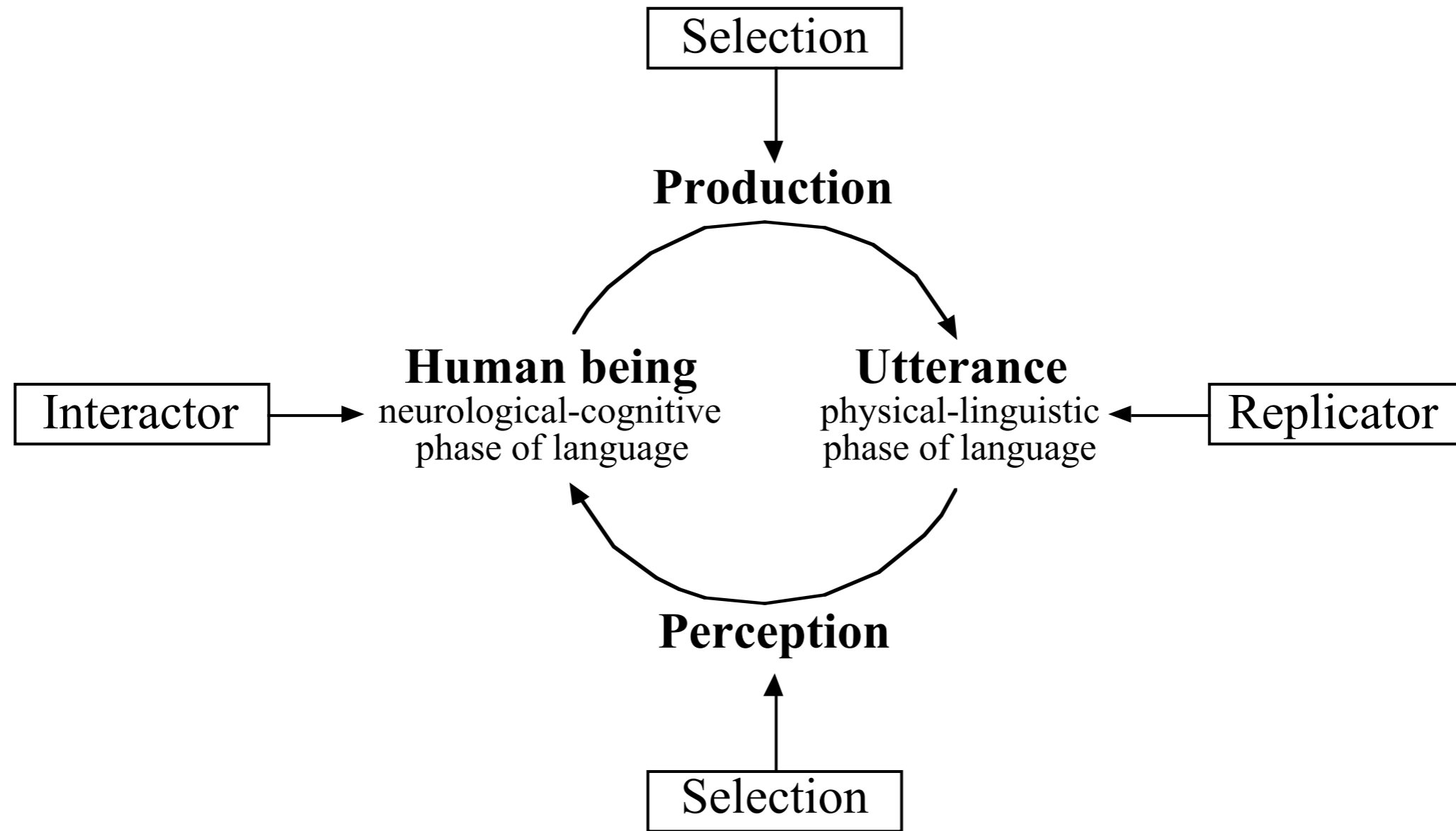
IDS	English	Hinukh
10.470	go	iλ'a
10.471	go up	šid iλ'a
2.330	marry	xoddoho yiλ'a
...

bezhta	0	1	0	0	1	0	0	0	0	1	1	1	0	1	0	0	0	1	0	0	1	1	0	0	0	0	0	1	1	
hinukh	0	1	0	0	0	1	0	0	0	0	1	1	1	1	0	0	1	1	1	0	1	1	1	0	1	1	0	1	0	0
hunzib	0	0	0	0	1	0	0	0	0	1	1	1	1	1	0	0	1	0	0	0	1	1	0	0	0	0	0	1	1	0
khvarshi inxokvari	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	1	0	0
khvarshi khvarshi	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0	0	0
tsez mokok	1	1	0	0	0	0	1	1	1	0	1	1	0	1	0	0	1	0	0	0	1	1	0	0	0	0	1	1	0	0
tsez sagadin	0	1	0	0	0	0	0	1	0	0	1	1	0	1	0	0	1	1	0	0	1	1	0	0	0	0	0	1	0	0



Theory of Utterance Selection

(Croft 2000)



- Dawkins, Richard. 1976. The Selfish Gene. Oxford: Oxford University Press.
- Hull, David L. 1988. Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science. Chicago: University of Chicago Press.
- Croft, William. 2000. Explaining Language Change: An Evolutionary Approach. Harlow: Longman.
- Cysouw, Michael. 2006. review of "Explaining Language Change," Beiträge zur Geschichte der deutschen Sprache und Literatur 128(2): 298-306.