

“PROTOLANGUAGES’ VS. LINGUISTIC NETWORKS ACROSS LANGUAGE BRANCHES’. A BASIC INVENTORY FOR RELATIONS OF CONCEPTS IN PREHISTORIC STATES OF LINGUISTIC COMMUNICATION

[Fee-Alexandra HAASE](#)

Professor, Ph.D. (University of Nizwa, Oman)

F.A.Haase@gmx.de, <https://orcid.org/0000-0003-4698-6955>

Abstract

The following is a study of the representation of the reflexes of the roots of the Borean language family in the Eurasiatic, Afroasiatic, Sino-Caucasian, Austric, and Amerindian language families. This Borean language family is considered to be a macro-language family, which comprises the aforementioned language families as their members. With the selection of words of this macro-family we can state that the roots of this macro-family are neither equally distributed among the member families nor distributed in the framework of the traditional borders of natural languages families. In order to check the distribution of these Borean roots, we use the “100-Words-Swadesh List” to check extensions of the related roots within the 5 different language families. The results are presented in a list of 100 concepts and a theory of the conditions of linguistic communication in pre-historic time, for which linguists usually employ the concept of ‘protolanguages’. We derive from our findings that the early hypothetical linguistic states of humans allowed the representation of oneself and the environment as its thesaurus entailed the conceptual meanings for it. On the contrary, a grammaticalization in the traditional understanding is not obviously resulting from the material as well as unlikely to have happened. The linguistic state was a ‘phonetic networking network’ across language families, so that the concept ‘language family’ should be critically revised.

Keywords: protolanguage, concept, a prehistoric state, linguistic, communication

Rezumat

Articolul înglobează un studiu amplu asupra protoformelor boreane în limbile eurasiatice, afroasiatice, sino-caucaziene, austrice și amerindiene. Prezența acestor protoforme în limbile date le permite unor savanți să identifice chiar o macro-familie boreană. Cercetarea ei demonstrează că protoformele boreane nu se înregistrează echilibrat în limbile menționate, iar, deseori, depășesc chiar granițele tradiționale ale acestei macro-familii. Pentru a investiga prezența și distribuția protoformelor boreane în limbile eurasiatice, afroasiatice, sino-caucaziene, austrice și amerindiene, folosim „Lista Swadesh de 100 de cuvinte”. Rezultatele studiului sunt prezentate într-o listă de 100 de concepte și o teorie a condițiilor comunicării glotice la etapa preistorică. Teoria dată, dezvoltată de o serie de lingviști, pune în uz conceptul de „protolimbă”. Investigația întreprinsă pe paginile acestui articol ne-a permis să conchidem că comunicarea prin „protolimbă” s-a făcut datorită reprezentării ca tezaur a sinelui și a mediului înconjurător. În limbile eurasiatice, afroasiatice, sino-caucaziene, austrice și amerindiene, protoformele boreane sunt de natură fonetică. Gramaticalizarea, în înțelesul ei tradițional, este puțin probabil să se fi produs. Aceste fapte îi determină pe unii cercetători să propună o revizuire a conceptului de „familie de limbi”.

Cuvinte-cheie: *protolimbă, concept, stare preistorică, lingvistică, comunicare*

1. Introduction: About the need of theory of linguistics at the end of linguistic history

The following article is an approach to find answers to the question of relations between languages transgressing the traditional classification systems of language branches of natural languages. While no doubt exists among linguists that natural languages change in a diachronic range for historically documented chronological ranges of natural languages, the situation is completely different for languages and/or linguistic varieties, which lack historical documentation, since they were just orally transmitted. Nevertheless, historical linguistic studies have since the 19th century discovered that the branches of language families have hypothetical common ancestors or at least share a common morphological material at the level of the words of their lexical inventory.

The contemporary state of knowledge in linguistics represents that the concept of 'language' is multilayered and also alternative conceptual frameworks should be taken into account in order to describe the linguistic situation for pre-historical linguistic states. The state of lack of historical material leads *nolens volens* to the necessity of theoretical frameworks about linguistic states in pre-historical times. The scientific approach needs to be experimentally based on non-empirical comparisons of documented linguistic material from different branches and the detection of the maximum amount of commonly shared linguistic features within the lexicon of the linguistic varieties and/or languages. This work of historical linguistic research has brought forward since the 19th century not only the knowledge of relationships between widely disperses languages as language families (which also tells us something about the migration backgrounds that can be additionally compared to findings in archeology and anthropology as supporting evidence). Also systematic linguistic inventory lists of the commonly shared lexical inventory have been presented by linguists. Pokorny's list of Indo-European roots as well as the Starostinian approach of multiple comparisons of similar material across language families supported by rich database material can be mentioned here besides selective studies in 'language contacts' (a problematic term we like to exchange against 'linguistic communication'), which uncover linguistic communication beyond the area of the commonly established language families.

The question how a community of speakers actually evolves increases our relative helpless state of methodological approaches, as even the theoretical material seem to be in need of a revision. The assumption that a language exists in pre-historical times, means that also the contemporary dominant associations of the concepts 'language' is applicable for pre-historic linguistic states. But this is not the case and we will in the following article based

on the material of the Borean language family demonstrate how the a community of speakers through differentiation of linguistic material and secondary grammaticalization could have communicated without the social and cultural borders we usually associate to linguistic varieties.

2. The state of research: language as faculty of linguistic communication, the concept 'protolanguage', and the idea of an evolution of languages

P. Kiparsky's article *Historical Linguistics and the Origin of Language* (Kiparsky, 1976, pp. 97–103) still refers to the idea of an origin of language in the tradition of the philologies of the 18th and 19th centuries. In *Reconsidering the "Isolating Protolanguage Hypothesis." in the Evolution of Morphology* J. Dubé wrote: "Much recent work on the evolution of language assumes explicitly or implicitly that the original language was without morphology. Under this assumption, morphology is merely a consequence of language use: affixal morphology is the result of the agglutination of free words, and morphophonemic (MP) alternations arise through the morphologization of once regular phonological processes (Dubé, 2013). This hypothesis is based on at least two questionable assumptions: first, that the methods and results of historical linguistics can provide a "window" on the evolution of language, and second, based on the claim that some languages have no morphology (the so-called isolating languages), that morphology is not a necessary part of language. The aim of this paper is to suggest that there is in fact no basis for what I will call the "Isolating Proto-Language Hypothesis" (henceforth IPH), either on historical or typological grounds, and that the evolution of morphology remains an interesting question". J. Dubé mentioned that "to appeal to a random genetic mutation or to a stroke of lightning in order to explain the evolutionary origins of language, or of one of its components, is clearly not an explanation, but it is equally unhelpful to keep a demonstrably wrong hypothesis because it may be the only available alternative to date to the argument from ignorance. Both of these views risk having the effect of preventing interesting investigations of the *problem* (not the *mystery*) of the evolution of morphology" (*ibidem*). K. R. Gibson in *Language or Protolanguage? A Review of the Ape Language Literature* described actual ape behaviors without prejudging their linguistic nature concluding "that a number of apes mastered essential components of protolanguage, but none constructed hierarchically structured sentences containing embedded phrases or clauses" (Gibson, 2012). J. R. Hurford in *The Evolution of Language and Languages* mentioned that "Bickerton's term *protolanguage* is a useful attention-focussing device, postulating that the class of 'languages' biologically available to *Homo erectus* was the class of protolanguages, defined quite roughly as systems for concatenating vocabulary with none of the complex syntactic dependencies, constituencies, command and control relations characterizing modern

languages. A *Homo erectus* individual, even if somehow presented with modern linguistic experience, could not make of it what a modern child makes of it, due to innate limitations" (Hurford, 1999). The researcher stated that "the language faculty has evolved as other genetically determined traits have evolved, via selection over the millions of alleles that contribute to the human genome. The phylogenetic evolution of the language faculty must have been slower by several orders of magnitude (assuming one could even quantify such things) than the sociocultural evolution of individual languages" [*ibidem*]. Proto-Indo-European is "presumed to have been spoken somewhere in Eastern Europe about five thousand years ago, and Proto-Iroquoian, the ancestor language from which the modern American languages of the Iroquoian family, such as Mohawk, are descended". J. R. Hurford (1999) and D. Bickerton (2014) wrote regarding the timing of protolanguage: "The question of when protolanguage emerged merits a brief consideration, if only because an adequate theory must eventually be able to integrate language evolution into the overall development of the genus *Homo*. The wide range of estimates in the literature (from the australopithecine era to the emergence of our own species) suggests that there are inadequate constraints to determine dates at this point. Much depends, too, on the nature of the initial selective pressure.

If this was scavenging (as suggested above) the likeliest time of onset would be between two and three million years ago. It seems intrinsically plausible that a longish period elapsed between the emergence of symbolic units and the emergence of syntax, since complex connections not required by any previous brain operations had to be forged, while perhaps also a critical mass of "recruitable neurons" had to be achieved".

The usual approach for the reconstruction of linguistic states, which lack historical documentation, is the reconstruction of a hypothetical former state based on the material of words in historical languages, which show similarities. These similar words can be cognates or other phonetically similar words. A common hypothetical ancestral root of the protolanguage both share is made in a method called 'comparative method'. The state of the protolanguages is discussed, as one position refers to them as hypothetical, whereas another position states that these protolanguages have existed and were used around 5000 BCE. In other words: The previously mentioned approach tends to an evolutionary perspective of languages including protolanguages going hand in hand with the evolution of the human from the apes, which already partly were able to produce and use features of the protolanguages (see Gibson, 2012). Al. Bouchard-Côté, D. Hall, Th. L. Griffiths, and D. Klein in *Automated Reconstruction of Ancient Languages using Probabilistic Models of Sound Change* (2013) stated that "one of the oldest problems in linguistics is reconstructing the words that appeared in the protolanguages from which modern languages evolved. Identifying the forms of these ancient languages makes it

possible to evaluate proposals about the nature of language change and to draw inferences about human history” (Bouchard-Côté *et alii*, 2013). According to these linguists, “protolanguages are typically reconstructed using a painstaking manual process known as the comparative method. We present a family of probabilistic models of sound change as well as algorithms for performing inference in these models. The resulting system automatically and accurately reconstructs protolanguages from modern languages” (*ibidem*). They stated too that “the key observation that makes reconstruction from these data possible is that languages seem to undergo a relatively limited set of regular sound changes, each applied to the entire vocabulary of a language at specific stages of its history” (*ibidem*). Like for other approaches of historical linguistics, the phonetic material is the subject of research: “Using phonological representations allows us to perform reconstruction and does not require us to assume that cognate sets have been fully resolved as a pre-processing step. Representing the words at each point in a phylogeny and having a model of how they change give a way of comparing different hypothesized cognate sets and hence inferring cognate sets automatically”. Al. Bouchard-Côté, D. Hall, Th. L. Griffiths, D. Klein use a probabilistic model of sound change and a Monte Carlo inference algorithm “to reconstruct the lexicon and phonology of protolanguages given a collection of cognate sets from modern languages. As in other recent work in computational historical linguistics (13–18), we make the simplifying assumption that each word evolves along the branches of a tree of languages, reflecting the languages’ phylogenetic relationships. We model the evolution of discrete sequences of phonemes, using a context-dependent probabilistic string transducer”. The database *Tower of Babel* entails indexes of the major language families with roots of protolanguages and the hypothetical Borean root as the antecedent root with representations in the main branches of language families. Based on the Borean root, it is possible to trace the relationships between words of protolanguages and the related reflexes in natural languages.

3. Question, methodology, and approach of our research: change and variations as underlying principles of development of ‘linguistic communication’

Our study is based on the corpus of the Tower of Babel project. In the tradition of ‘big questions’ of humankind the question how words and things refer to each other would be the guiding question of our research. In academia the question is approached since antiquity by philosophers and rhetoricians and in the heritage of the tradition of Western academic studies in linguistics, semiotics, and communication sciences further answered. How linguists interpret the findings from the comparisons of linguistic varieties varies. While naïve early 19th century researchers in the tradition of Herders *Ursprache* still assumed the existence of arch-languages as ground for later languages, the framework of the concept ‘language’ was more and more critically revised. Still the term ‘proto-language’ is commonly used for a hypo-

thetical antecedent language/linguistic variety. The simplest description for the common features we are interested in is an inductively derived morphological material at word-level, which is shared among language varieties. Our research interest focuses on the common linguistic features of languages beyond the level of a single branch of languages. The Starostinian database and the concept of the Borean macro-language family allow us to have material for the comparison of material beyond the level of language families. The term 'linguistic communication' is used for any linguistically relevant material, which also entails our hypothetical material as representations of 'linguistic communication'.

As propedeutics for the problem and our discussion and as the description of the segment of our research in a case study we select a semantic framework idea of most common words of a linguistic variety and semantic categories of lexical material. It is statistically possible to rank the most commonly used words within a text corpus of a thesaurus. An example is the list of the most commonly used words in the English language (see table below). On the contrary, it is not possible to determinate the exact meaning of words, when operating at the level of hypothetical linguistic states, since here *conceptual meanings* are associated to the material. We also have here examples of semantic framework ideas, which aim at the classification of the inventory of a linguistic variety according to the closeness to this field. Usually, these categorizations refer as *orbis pictus* to the world. The *Semantic Fields for Indo-European Linguistics* used at the University of Texas at Austin for the categorization of semantic fields of Pokorny's etyma (see table below). The semantics of the roots of the Borean hypothetical ancestral linguistic variety result from the semantic meanings of the Starostian databases at *Tower of Babel*. (Below we discuss their actual function as concepts rather than lexicological meanings). For our theoretical framework, we rely on the theory of transpositional grammar based on the assumption that parts of speech can be subject of a transposition from one to another part of speech and also complex expressions can be associated to this process.

J. R. Hurford stated that "historical linguists have catalogued many types of change that can occur in the evolution of individual languages, changes such as weakening and strengthening of the meanings of words, change of basic word order, loss of inflections, grammaticalization of lexical words (nouns, verbs, adjectives) into grammatical function words (articles, pronouns, auxiliaries), merger of phonemes, the emergence of novel phonemic distinctions, lowering, raising, fronting, backing and rounding of vowels, palatalization, glottalization, and so on" (Hurford, 1999). The selected 100 words from the *Swedish List* are analyzed regarding their individual reflexes in the main language families, which form cognates across language families. The sets of associated reflexes in protolanguages are the *nodi*, which connect the conceptual Borean roots with natural languages across the main lan-

guage families. Concerning the main aspects of linguistic studies, the historical linguists rely on a symbolically recording system of phonetic sounds for words and reconstructed or evident roots, which is usually IPA. As the symbolic representations are strings of phonetic values, the morphological aspect of the lexicon is also present. When it comes to semantics, as soon as we leave the area of the historical records, the associated meaning must be considered to be a conceptual meaning, since grammaticalization and differentialization are not reconstructable in linguistic states prior to the existence of historical natural languages. So the meaning 'green' is actually the concept 'green' allowing grammatical alterations like 'greenish', 'green', 'the green', and 'be green' usually coming together with a grammaticalization of the roots. About the syntactical formation and transposition of the lexical inventory of protolanguages we know less. Usually word orders in word compounds (like sentences) and change of the root or its alteration with grammatical markings of single words are the syntactical features of natural formats of linguistic communication like natural languages.

4. Propedeutics of contrastive studies: word lists and other semantic-driven linguistic inventories

A hypothetical construct of a lexical thesaurus of a community of speakers is produced by linguists with top-used lists of words of thesauri of natural languages. For a community of speakers this hypothetical construct would represent the most commonly used words shared in linguistic communication acts. The most commonly used words for the community, which used the roots of the communication system of the Borean roots of course cannot be determine. The actual problem of the semantic meanings associated to roots will be discussed below and instead of them, the concept, from which the transpositions derive, will be promoted as theoretical approach. In contrast to the *Swedish List*, contemporary top-100-lists entail grammatical and lexical elements, which are linguistic and grammatical markers (like below *the* and *a*, *be*, *of*, *if*, *her* and *she* and *his* and *him*, *would* and *could* as conditionalizing markers, prepositions like *at* or *in*, *which*, *when*, *or* and *and*, *not*, *just* and *only*, *because* and *so*, which are not referring to the semantic world of things, but their relations in a linguistic framework. They are usually from natural language to natural language differently expressed. In the *Swedish List* they don't exist. According to the *Oxford English Corpus* (2014), the following are the most used words in the English language as of the year 2014:

Rank	Word	Rank	Word	Rank	Word	Rank	Word	Rank	Word
1	<i>The</i>	21	<i>This</i>	41	<i>So</i>	61	<i>people</i>	81	<i>Back</i>
2	<i>Be</i>	22	<i>But</i>	42	<i>Up</i>	62	<i>Into</i>	82	<i>After</i>
3	<i>To</i>	23	<i>His</i>	43	<i>Out</i>	63	<i>Year</i>	83	<i>Use</i>
4	<i>Of</i>	24	<i>By</i>	44	<i>If</i>	64	<i>Your</i>	84	<i>Two</i>
5	<i>And</i>	25	<i>From</i>	45	<i>About</i>	65	<i>Good</i>	85	<i>How</i>
6	<i>A</i>	26	<i>They</i>	46	<i>Who</i>	66	<i>Some</i>	86	<i>Our</i>

7	<i>In</i>	27	<i>We</i>	47	<i>Get</i>	67	<i>Could</i>	87	<i>Work</i>
8	<i>That</i>	28	<i>Say</i>	48	<i>Which</i>	68	<i>Them</i>	88	<i>First</i>
9	<i>Have</i>	29	<i>Her</i>	49	<i>Go</i>	69	<i>See</i>	89	<i>Well</i>
10	<i>I</i>	30	<i>She</i>	50	<i>Me</i>	70	<i>Other</i>	90	<i>Way</i>
11	<i>It</i>	31	<i>Or</i>	51	<i>When</i>	71	<i>Than</i>	91	<i>Even</i>
12	<i>For</i>	32	<i>An</i>	52	<i>Make</i>	72	<i>Then</i>	92	<i>New</i>
13	<i>Not</i>	33	<i>Will</i>	53	<i>Can</i>	73	<i>Now</i>	93	<i>Want</i>
14	<i>On</i>	34	<i>My</i>	54	<i>Like</i>	74	<i>Look</i>	94	<i>Because</i>
15	<i>With</i>	35	<i>One</i>	55	<i>Time</i>	75	<i>Only</i>	95	<i>Any</i>
16	<i>He</i>	36	<i>All</i>	56	<i>No</i>	76	<i>Come</i>	96	<i>These</i>
17	<i>As</i>	37	<i>Would</i>	57	<i>Just</i>	77	<i>Its</i>	97	<i>Give</i>
18	<i>You</i>	38	<i>There</i>	58	<i>Him</i>	78	<i>Over</i>	98	<i>Day</i>
19	<i>Do</i>	39	<i>Their</i>	59	<i>Know</i>	79	<i>Think</i>	99	<i>Most</i>
20	<i>At</i>	40	<i>What</i>	60	<i>Take</i>	80	<i>Also</i>	100	<i>Us</i>

Table 1: *The most used words in English as of the year 2014*

The contemporary text corpus-based 100 most common words in English can – in contrast to the conceptual semantic meanings of proto-material - be classified with exactly one part of speech. In hypothetical linguistic communication material the meaning is conceptual and thus beyond the level of grammaticality. According to the parts of speech the ranking of the top-100 words in English is as follows:

Nouns (25)	Verbs (25)	Adjectives (25)	Prepositions (17)	Others (25)
<i>time</i>	<i>be</i>	<i>good</i>	<i>to</i>	<i>the</i>
<i>person</i>	<i>have</i>	<i>new</i>	<i>of</i>	<i>and</i>
<i>year</i>	<i>do</i>	<i>first</i>	<i>in</i>	<i>a</i>
<i>way</i>	<i>say</i>	<i>last</i>	<i>for</i>	<i>that</i>
<i>day</i>	<i>get</i>	<i>long</i>	<i>on</i>	<i>I</i>
<i>thing</i>	<i>make</i>	<i>great</i>	<i>with</i>	<i>it</i>
<i>man</i>	<i>go</i>	<i>little</i>	<i>at</i>	<i>not</i>
<i>world</i>	<i>know</i>	<i>own</i>	<i>by</i>	<i>he</i>
<i>life</i>	<i>take</i>	<i>other</i>	<i>from</i>	<i>as</i>
<i>hand</i>	<i>see</i>	<i>old</i>	<i>up</i>	<i>you</i>
<i>part</i>	<i>come</i>	<i>right</i>	<i>about</i>	<i>this</i>
<i>child</i>	<i>think</i>	<i>big</i>	<i>into</i>	<i>but</i>
<i>eye</i>	<i>look</i>	<i>high</i>	<i>over</i>	<i>his</i>
<i>woman</i>	<i>want</i>	<i>different</i>	<i>after</i>	<i>they</i>
<i>place</i>	<i>give</i>	<i>small</i>	<i>beneath</i>	<i>her</i>
<i>work</i>	<i>use</i>	<i>large</i>	<i>under</i>	<i>she</i>
<i>week</i>	<i>find</i>	<i>next</i>	<i>above</i>	<i>or</i>
<i>case</i>	<i>tell</i>	<i>early</i>		<i>an</i>
<i>point</i>	<i>ask</i>	<i>young</i>		<i>will</i>
<i>government</i>	<i>work</i>	<i>important</i>		<i>my</i>
<i>company</i>	<i>seem</i>	<i>few</i>		<i>one</i>
<i>number</i>	<i>feel</i>	<i>public</i>		<i>all</i>
<i>group</i>	<i>try</i>	<i>bad</i>		<i>would</i>

<i>problem</i> <i>fact</i>	<i>leave</i> <i>call</i>	<i>same</i> <i>able</i>		<i>there</i> <i>their</i>
-------------------------------	-----------------------------	----------------------------	--	------------------------------

Table 2: The ranking of the top-100 words in English

G. Starostin in *Preliminary Lexicostatistics as a Basis for Language Classification: A New Approach* presented a 50-item-wordlist for the global lexicostatistical database (Starostin, 2014). This list entails all words of the Swadish List, but ranks them differently:

<i>we</i>	<i>moon</i>	<i>sun</i>
<i>two</i>	<i>finger nail</i>	<i>smoke</i>
<i>I</i>	<i>blood</i>	<i>tree</i>
<i>eye</i>	<i>one</i>	<i>ashes</i>
<i>thou</i>	<i>tooth</i>	<i>rain</i>
<i>who</i>	<i>new</i>	<i>star</i>
<i>fire</i>	<i>dry (e.g. of clothes)</i>	<i>leaf</i>
<i>tongue</i>	<i>eat</i>	<i>kill</i>
<i>stone</i>	<i>tail</i>	<i>foot</i>
<i>name</i>	<i>hair (of head)</i>	<i>horn</i>
<i>hand</i>	<i>water</i>	<i>hear</i>
<i>what</i>	<i>nose</i>	<i>meat (as food)</i>
<i>die</i>	<i>not</i>	<i>egg</i>
<i>heart</i>	<i>mouth</i>	<i>black</i>
<i>drink</i>	<i>ear</i>	<i>head</i>
<i>dog</i>	<i>bird</i>	<i>night</i>
<i>louse (head)</i>	<i>bone</i>	

Table 3: G. Starostin's 50-Words List (2014)

Did a community of speakers have the same 'orbis pictus' have in mind and words like a contemporary human? Definitely not, but some basic fields like the nature or human features are continuingly stable. A list of semantic fields of the lexicon of languages was made in the Department of Linguistics of the University of Texas at Austin (IEL). The Department of Linguistics of the University of Texas at Austin employs the following semantic fields for the historical linguistic thesaurus:

Agriculture/Veg.	Motion/Transport
Animals	Physical Acts/Mat'ls
Body Part/Function	Physical World
Clothing/Adornment	Possession/Trade
Dwellings/Furniture	Quantity/Number
Emotion	Religion/Beliefs
Food/Drink	Sense Perception
Language/Music	Social Relations
Law/Judgment	Spatial Relations
Mankind	Time

Mind/Thought	Warfare/Hunting
--------------	-----------------

Table 4: *Indo-European Linguistics. Semantic Fields. List 1. University of Texas at Austin*

The Department of Linguistics of the University of Texas at Austin employs another list of semantic fields for the historical linguistic thesaurus:

Physical World	Spatial Relations
Mankind	Quantity & Number
Animals	Time
Body Parts & Functions	Sense Perception
Food & Drink	Emotion
Clothing & Adornment	Mind & Thought
Dwellings & Furniture	Language & Music
Agriculture & Vegetation	Social Relations
Physical Acts & Materials	Warfare & Hunting
Motion & Transportation	Law & Judgment
Possession & Trade	Religion & Beliefs

Table 5: *Indo-European Linguistics. Semantic Fields. List 2. University of Texas at Austin*

The classifying categories for the ‘*orbis pictus*’ or – otherwise expressed – the thesaurus of a human linguistic communication enabling system (like a natural language) usually aim at presenting an abstract reference framework for a thesaurus. The *Swedish List* is simpler than the categories of the semantic fields and lacks any complex conditions of single semantic items. Nearly all of the semantic meanings or words of the *Swedish List* are present in the database *Tower of Babel* in the Borean-Proto-Level Database. Our study is based on the so-called ‘words’ of the *Swedish List*. The *Swadesh List* is a compilation of basic words for the purposes of historical-comparative linguistics (S100WL, 2014). This choice derives from the needs of a semantic selection than from the actual indexing of this list. We discuss the theoretical problems of the *Swedish List* at another place of this article. The list entails the following entries:

<i>I</i>	<i>horn</i>	<i>lie</i>
<i>You</i>	<i>tail</i>	<i>sit</i>
<i>we</i>	<i>feather</i>	<i>stand</i>
<i>this</i>	<i>hair</i>	<i>give</i>
<i>that</i>	<i>head</i>	<i>say</i>
<i>who?</i>	<i>ear</i>	<i>sun</i>
<i>what?</i>	<i>eye</i>	<i>moon</i>
<i>not</i>	<i>nose</i>	<i>star</i>
<i>all</i>	<i>mouth</i>	<i>water</i>
<i>many</i>	<i>tooth</i>	<i>rain</i>
<i>one</i>	<i>tongue</i>	<i>stone</i>
<i>two</i>	<i>claw</i>	<i>sand</i>

big	foot	earth
long	knee	cloud
small	hand	smoke
woman	belly	fire
man	neck	ash(es)
person	breasts	burn
fish	heart	path
bird	liver	mountain
dog	drink	red
louse	eat	green
tree	bite	yellow
seed	see	white
leaf	hear	black
root	know	night
bark	sleep	hot
skin	die	cold
flesh	kill	full
blood	swim	new
bone	fly	good
grease	walk	round
egg	come	dry
		name

Table 6: *Swadesh's 100-Word List. Department of Linguistics. Portland State University*

5. The interpretation of findings: 'how was pre-historic linguistic communication like?'

Even though we rely in research on the database for the so-called Borean proto-language, we are still aware that the concept 'language' is here more than problematic and we will finally discuss alternatives below. Nevertheless, the merits of the database and the approach of Starosin are evident. The linguistic communicative macro-format we suggest for this kind of linguistic communication is out of the range of the natural language as far as we describe it within the framework of our linguistic method; the list of the Borean roots, which follow the pattern *K1-V2-K1-V2* for *consonant 1 – vowel 1 – consonant 2 – vowel 1* within the traditional root scheme of morphology of words, refers to a general word pattern of two radical consonants followed by a vowel. The Borean roots of one of the semantic meanings refer to at least one of the following language families:

- E: Indo-European Language Family
- A: Afroasiatic Language Family
- SC: Sino-Caucasian Language Family
- AU: Austric Language Family
- AM: Amerindian Language Family

The hypothetical Borean roots, our material from the *Tower of Babel* database, we indexed according to the 100 top common semantic fields of the

Swadish List (see table below). Generally speaking, we can say that the majority of the semantic fields of the *Swadish List* has more than one reference to Borean roots. The majority of the Borean roots also refer to more than one language family. The presence of multiple Borean roots in one language family for the representation of a linguistic concept can be interpreted as:

1. Presence of Multiple Borean roots and their reflexes in one language family;
2. Indicator for a sub-segmented distribution of the Borean roots and their reflexes in the topographical area associated to a language family;
3. The general occurrence of root material in more than one of the traditional language families.

Some of the semantic fields of the *Swadish List* are not or surprisingly thin covered by Borean roots (e.g. 'drink' and some colors). This can be interpreted as a lack of the coverage of these semantic fields in the text corpus of the database *Tower of Babel*. The data derived from this database are probably not sufficient for the coverage of all existing roots (below we discuss this as phenomenon of transpositional grammar). We assume that the semantic fields actually allow multiple grammatical realizations in the sense of the transposition grammar, which was in disciplines like Egyptology successfully employed for the description of linguistic features of an early language, which evolved within the transformation of images to writing. The here presented roots are concepts, which have either no concrete part of speech as associated grammatical feature and/or we are not able to construct grammatical features like the associated part of speech out of the material. On the one hand we have an extreme interwoven pattern of concepts and multiple roots across several language families, or the other hand we have a low level of grammatical features. Except the pronouns, we have a morphological pattern of two radicals (Ancient Egyptian has as standard pattern three radicals). This material allowed alterations for the purpose of grammatical features and in this regard the structure would not be very different from contemporary languages, which use additional sounds for grammatical markings).

6. An outlook to a linguistic categorization: towards a theory of concepts and conceptualization of (pre-historic) linguistic communication as 'phonetic networking network'

In the tradition of transposition grammar we conclude that the semantic inventory of the *Swadish List* must be in the context of our root analysis considered the conceptual inventory of a thesaurus, which allows transpositions of parts of speech for each concept. Alternatively, it can also be assumed that the morphological material was simply not semantically differentiated; it was applied and within the course of the usage among speakers individual meanings as well as the differentiation of grammatical features occurred and was coded within a selected community of speakers. In the theory of 'trans-

position grammar' the generation of parts of speech is according to transposition grammatical approaches a secondary phenomenon, as an expression can be transposed from one part of speech into another one generating syntactical patterns. The actual appearance of the semantic connotation associated to the roots must be classified as conceptual meaning or concept lacking any grammatical aspects. The meanings are grammar-free concepts, which can be transformed (when realized) into grammatical forms. This conceptuality of the meanings of the roots has to be critically mentioned, when commenting the associated meanings.

(1) Nominal Expressions	Humans	Nouns and Nominal Expressions
	Objects	Nouns and Nominal Expressions
(2) Verbal Expressions	Actions	Verbs and Verbal Expressions
(3) Adjectival Expressions	Nominal Qualities	Adjectives and Adjectival Expressions
(4) Adverbial Expressions	Verbal Qualities	Adverbs and Adverbial Expressions

Table 7: *Transpositional grammatical framework for lexicological entries*

In the following chart the types of actual expressions associated to the words of the *Swedish List* are added. Under the paradigm of the transpositional grammar the concept derive from each of the words is able to have its reflexes in the full range of expressions derived from the main parts of speech. Of course, the realization of the reflexes doesn't occur in all languages and the concept can be expressed in a certain part of speech with the use of another concept and roots:

I	NE	nose	NE	smoke	NE
You	NE	mouth	NE	fire	NE
we	NE	tooth	NE	ash(es)	NE
this	NE	tongue	NE	burn	VE
that	NE	claw	NE	path	NE
who?	NE	foot	NE	mountain	NE
what?	NE	knee	NE	red	ADJE
not		hand	NE	green	ADJE
all	NE	belly	NE	yellow	ADJE
many	NE	neck	NE	white	ADJE
one	ADJE	breasts	NE	black	ADJE
two	ADJE	heart	NE	night	NE
big	ADJE	liver	NE	hot	ADJE
long	ADJE	drink	VE	cold	ADJE
small	ADJE	horn	NE	lie	VE
		tail	NE		VE

woman	NE	feather	NE	sit	VE
man	NE	hair	NE	stand	VE
person	NE	head	NE	give	VE
fish	NE	ear	NE	say	NE
bird	NE	eye	NE	sun	NE
dog	NE	eat	VE	moon	NE
louse	NE	bite	VE	star	NE
tree	NE	see	VE	water	NE
seed	NE	hear	VE	rain	NE
leaf	NE	know	VE	stone	NE
root	NE	sleep	VE	sand	NE
bark	NE	die	VE	earth	NE
skin	NE	kill	VE	cloud	ADJE
flesh	NE	swim	VE	full	ADJE
blood	NE	fly	VE	new	ADJE
bone	NE	walk	VE	good	ADJE
grease	NE	come	VE	round	ADJE
egg	NE			dry	NE
				name	

Table 8: *Swadesh's 100-Word List. Department of Linguistics. Portland State University*

The question of the most important words in a thesaurus is – to conclude finally – an unnecessary question; as the transposition of words allows as derivations from one concept multiple realization of reflexes, this question and its answer can only be approached in limited empirical studies. One of the explanations of the multiple appearances of several roots related to different concepts within the Borean roots we examined is that the word lists like the *Swedish List* do not take into account that it should be conceptual meanings and not thesaurus inventory of languages. One root concept (SUN) could for example have represented an adjectival expression in one reflex (sunny), while another one was used for the noun 'sun' or 'shine'. The benefit of the attached list of Borean roots and their conceptual meaning lies in the representation of the widest range of trans-familiar linguistic relations for 100 common words.

The knowledge about pre-historic linguistic states of course needs to be similarly approached like historic linguistic phenomena and thus aims at answers for questions about the lexical, semantic, morphological, and phonetic qualities of representations of linguistic communication as reconstructed forms from historical documentation of historical linguistic material of natural languages. At the level of the concept, e.g. SUN, such linguistic features are irrelevant. The linguistic features become relevant as soon as the linguistic material is the topic of the research; in our case six Borean roots for the concept SUN exist; five of them are in more than one language family present.

NVJV	E/AF	Sun
KVMV	E/SC	Sun
HVKV	SC/AM	Sun
NVRV	E/AU	Sun
PVCV	AF	Sun
TVNV	E/SF/AU	Sun

The associated meaning 'sun' is actually a hypothetical or conceptual meaning. Approached by the transpositional grammatical theory, it could also refer to such lexical forms like shiny, shine (like the sun), shining etc. Our findings clearly support the existence of a relation of one word of the Swadish list to more than one semantic meaning associated to a root. This is a network, which is not governed by the rules of traditional language families. On the contrary, it exists as not by the borders of language families ruled network. As we are here in the hypothetical area of reconstruction, we have to face also the question of the qualities and features of the pre-historic state of linguistic communication. The problematic state of the term 'proto-language' was raised in recent research. The concept 'language' is here not sufficient, as it implicitly involves the idea that a natural language-like phenomenon exists. The material discussed here actually forces us to assume the contrary: a topographically diverse networked layer of conceptually related linguistic text elements at the micro-level of words. Factors for their extensions and limitations must be seeks in non-linguistic and/or socio-linguistic aspects. As for now, we formulate of theses as follows for global aspects of pre-historic linguistic communication:

1. Linguistic communication in pre-historic settings was not (only) ruled by the differentiation of languages in language evolution as we know it typically in sets of language families;
2. Micro-linguistic units like the examined roots demonstrate that words or other micro-linguistic contents of linguistic communication in their morphological material and the semantic meaning transcended borders of traditional language families;
3. Instead of the assumption of a word and a semantic meaning we must operate with concepts and conceptual meanings for the description of pre-historic states of linguistic communication.

However linguistic communication took part, we assume that it was less limited and structured as contemporary definitions might imply. The content of these pre-historic linguistic communication was highly communicable and less regulated by socio-cultural factors. Also its grammaticality stands at disposal. We know that older languages have a low level of grammatical markers. The exchange of vocals as the principle of grammaticalization is a feature of languages, which usually fall from a language-evolutionary perspective in a time frame, which was before the use of distinct grammatical markers (Arabic is a language, which still uses the vowel changing, while

in Germanic languages this principle was exchanged by implemented markers, which were/are not vowels). The relative simple sound patterns of the roots of the Borean language family allowed alterations among the communities of their users. We can assume that among the variety of roots offered the members of open speech communities saw themselves permanently in the position of selecting and specifying the linguistic communication material they received. Within this process actually a 'community of speakers' could rise and evolve and distinguish itself from the other users of linguistic communication.

Our interest is not to find evidence or contribute to a theory of the evolution of language as a human faculty or as a system of linguistic communication. The evolutionary approaches we discussed in the introductory part of this article assume that language evolves from simpler forms (like proto-languages) to higher patterns of linguistic communication like the contemporary natural languages. But this is not the aim of our studies here, which look for an answer regarding the kind of formation of language varieties. As far as we see the distribution of Borean roots, we can assume that linguistic communication operated, generally speaking, in all directions as a networking process of the networked substance for the communicative performance. This undirected and unstructured process was on the other hand modified by social and extrinsic (e.g. natural) conditions. The participation in a common phonetic symbolic system allowed participating in social activities. The stabilization or codification of the 'phonetic networking network' was not practiced and when it occurred, it marked in a region with a community of speakers the step from a-historical to historical time. We should not project our - actually accrued - contemporary understanding about language etc. onto the situation of the state of 'linguistic communication' before the rise of historical languages. We are able to reconstruct common aspects of groups of languages in the linguistic symbolic representation of phonetics. But we are not able to differentiate a grammatical structure. Even the association of a grammaticalization of former linguistic states is wrong, as the earliest natural languages were coded signs (China, Egypt). Letters as arbitrary signs did not exist, they developed later. Additional information, for which we usually use grammar (like plural, tense etc.), were and are implemented by markers for the change of a word or a specific grammatical phenomenon or the establishment of certain positions or orders of words. What we can say about the status of the linguistic communication is that specific patterns ('strings') of vowels and consonants were stable components of communication across generations in terms of their phonetic components and basic morphological components as well as concerning their reference to a specific semantic concept. Such strings for specific concepts existed in wide areas across the globe. They do not show the regional limitation to a certain region

like we find it in the separation of regions associated to the main branches of language families for historical languages. The communication of the humans able to perform linguistic communication was purely applied and allowed the sharing entities to communicate within a symbolic representative system of sounds the world around them and themselves as a semiotic process for references. We do not know if the communication at this state already entailed functional pragmatic aspects like expressions of orders, conditions, negations etc. But looking at the variety of linguistic forms for such functional aspects in natural languages, it is likely that this was a state of differentiation among smaller units of speakers driven by the rules of their own socialization.

Appendix: The Historical Linguistic Paths of the Semantic Fields of Borean Roots

Semantic value / concept	Annotated Borean root	The language families, in which reflexes of the respective Borean root exist
<i>I</i>	HVKV MV NV WV HVCV CV	E/A/SC/AU/AM E/A/SC/AU/AM E/A/SC/AU/AM E/A/AU A/SC/AU/AM SC/AU/AM
<i>You</i>	HU MV NV TV WV CV	E/A/SC/AM AU/AM E/A/SC E/AM/SC/AM SC/AM E/SC/AU/AM
<i>we</i>	PV TV LV WV	AU/AM SC/AU/AM SC SC/AU/AM
<i>this</i>	NV KV HV CV	E/A/SC/AU/AM E/A/SC/AM E/AF/SC/AM E/A/SC/AU/AM
<i>that</i>	MV PV TV HV(1) HV (2) HV (3)	A/SC/AM E/A/SC/AM E/A/SC/AU/AM E/A/SC/AM E/A/SC/AU/AM E/A/SC/AU/AM
<i>who?</i>	KV NV CV	E/A/SC/AU/AM E/A/SC/AU/AM E/SC/AU/AM

<i>what?</i>	MV PV RV	E/A/SC/AM/AU SC/AU/AM AU
<i>not</i>	CV HV HVLV HVN MV PV TV	A/SC/AU E/AF/SC/AU E/A E/A/AU E/A/SC A/SC SC/AU
<i>all</i>	HVLV WVNCV KVLV KVLWV	SC/AU/AM E/A/SC E/A E/A/SC/AM
<i>many</i>	PVLV MVNV PVHV PVTV	E/AF/SC/AM/AU E/A/SC/AM E/A/SC/AU/AM E/A/SC
<i>one</i>	HVRV TVKV HVTV HVN	E E/A/SC . AU/AM E/A E/AU
<i>two</i>	HVNLV CVRV JVRU MVLV TVWV CVNV	E/SC/AU A E E/A/SC E/A/SC/AM/AU A/SC/AU
<i>big</i>	HVN MVKV TVTV	E/A/SC E/A/AU/AM E/A/SC
<i>long</i>	KVLV	E/AM
<i>small</i>	KVRCV KVTV PVTV TVHV TVHV TVNV CVKV	SC/AU E/A/SC/AM E/A/SC E/A/SC/AU E/A/SC/AU E/SC/AM E/A/SC
<i>woman</i>	KVNV NVCV PVNV NVTV CVCV	E/A/SC/AU/AM E/A/SC E/SC/AU/AM E/A E/SC
<i>man</i>	KVNTV MVNV	E/A/SC/AM E/A/SC/AU/AM

	MVRV KVCV	E/A/SC/AU SC/AM
<i>person</i>	CVPV MVCV	E/A SC/AU/AM
<i>fish</i>	KVLV KVMVC KVRV LVMV NVNV TVKV CVMV	E/A/AM E/A/SC/AM E/A E/A/AM E/A E/A/SC E/A/AM
<i>bird</i>	PVHV KVLV KVMV KVVV KVPV KVRV KTVV LVKV SVKV TVRV CVKV CVPV HVWV	A/SC E/E/SC E/A E/SC/AU E/A E/A E/A/SC/AU E/A/SC/AU E/AM E/A E/A/SC/AM E/A E/A
<i>dog</i>	KTVV KVPV NVKV KVVV PVRV NVNTV CVKV CVRPV HVCV HVMV KVCV KVLV	A/AU E E/AM E/A/SC/AU/AM E/A/SC/AM E/A SC A/SC A/AU/AM A E/A/SC/AM A/AU
<i>louse</i>	MVHV NVJV TVJV JVLKV	AU E/A/SC E/A/SC/AU/AM A
<i>tree</i>	HVJWV HVLNV KVJWV KVRV NVKV PVJV	E/M E/A/SC/AU E/SC/AU/AM E/A/SC E/AM E/A/SC

	PVNCV TVNV TVRV CVLV WVTV	E/SC E/A/SC E/A/SC E E/A/AM
<i>seed</i>	WVTV	E/A/AM
<i>leaf</i>	PVLV HVRLV LVPV PVNKV	E/A/SC E/A/SC/AU E/A/SC E/SC/AU/AM
<i>root</i>	TVMV MVRKV KVRV WVRTV	E/A/SC E/A/SV A/SC E/SC
<i>bark</i>	KVRV KVLV KVNKV KVRPV PVKV CVPV KVPV	E/A/SC/AU E/E/SC E/A/SC E/A/SC E/A/SC/AU E/A/SC E/A/SC/AU
<i>skin</i>	KVLV	E/A/SC/AU
<i>flesh</i>	TVLV HVMCV NVKRV CVLV CVCV CVKV	SC E/SC/AM E/SC/AU E/SC E/A/SC/AU E/A/SC
<i>blood</i>	CVLV PVLV KVRV WVRV HVNV KVN PVRV CVHV CVMV CVNV	E / SC E / A / AM E / SC E / SC / AU E / SC / AU E A E / SC / AM A E / A
<i>bone</i>	TVKV PVNV KVRV LVNV MVKV RVNV CVMV CVNV	AU E/SC/AM A SC/AU A/SC/AM SC/AU E/A/SC/AU/AM E/SC

	HVCV HVKV HVMTV KVCV	E/A/SC/AU/AM A/AM SC E/AF/SC/AM
<i>grease</i>	KVJV CVPV MVRV CVMV	E/SC/AU/ AM E / AF / SC E / AF / SC / AU AF / E / SC / AU
<i>egg</i>	LVNV TVLV	SC / AU E / SC
<i>horn</i>	MVRKV LVRV KVRV	E/AF/SC SC E/AF/SC
<i>tail</i>	PVNCV CVPV CVRV KVRV KVTV MVHV TVKV TVNV	E/SC/AU AF E/AF AF/AU E/AF/SC/AU/AM SC E/AM AU
<i>feather</i>	KVMV PVLV	AF/SC/AM E/AF/SC/AU/AM
<i>hair</i>	PVWV KVRV PVTV KVLV NVJV NVNV NVRV PVLVCV PVNV PVRV TVKV TVMV TVRV CVKV CVRV KVCV LVHV	SC/AU/AM E/SC E/SC/AU/AM E/AF/SC E/AF/SC E/SC E SC E E/AF/SC E/AF/SC/AM E/AF E E/AF/AC/AU E/SC/AU E/SC/AU AF/SC/AM
<i>head</i>	KVHV MVHV MVTV WVNLV WVTMV CVKV	AF/SC/AU/AM E/SC E/AF/SC/AU SC/AU E/SC SC/EM

<i>ear</i>	KVRV LVLV MVHV MVNCV CVRV HVNv	E/AF/ AU SC/ AU E E/SC AU SC
<i>eye</i>	HVNv HVPV WVMKV HVKV CVLV	E/AF/SC/AU/ AM AF/AM SC E/AF/SC/AU/ AM E
<i>nose</i>	NVKCV NVRV TVNV WVKV CVRV KVNv	E/AF/AM E/AF/SC E/AF/AU E E/AF E/SC
<i>mouth</i>	HVPV HVWV KVMPV	E/AF/ AU/AM E/AF E/SC/ AU
<i>tooth</i>	HVNv LVKV MVNV PVLV	E/SC/AU A/AF/SC/ AU AF/SC AU/E
<i>tongue</i>	MVLV KVLV LVLV NVNV SVMV SVWV CVHV	SC/AU/AM E/AF/SC/AM E/SC/AU E/AU/AM AU SC SC
<i>claw</i>	-	-
<i>foot</i>	PVLMV KVLV HVPV KVNv KVRV KVSv LVKV MVLV PVCKV CVKV	AF/SC/AM SC/AU E/SC SC/AU E/SC SC/AM E/AF/SC SC E/AF/SC/AM E/AF/SC/AM
<i>knee</i>	-	-
<i>hand</i>	LVLV KVNv CVPV	E/SC/AU/AM E/AF/SC/AM E/AF

	KVRV MVNV PVCV PVKV PUNKV	E/AF E/AF/SC/AU/AM AF/SC E/AF/SC/AM E/SC/AU
<i>belly</i>	KVRPV KVTV PVHV PVNKV PVRV TVKV WVNCV KVVV HVLV	E/AF/SC E/AF/SC/AM AF/SC/AU/AM SC/AU/AM E/AF SC/AU/AM E/AF/SC/AM E/AF/SC/AM E/AF/SC/AM
<i>neck</i>	KVLV KVNKV	E/AF/SC/AM E/AF/SC
<i>breasts</i>	KVRTV MVLKV MVNV WVNXV	E/AF/SC E/AF/SC/AM E/AF/SC/AU AF/SC/AU
<i>heart</i>	RVKV CVLMV	SC/AU E
<i>liver</i>	TVRNV CVNCV KVLV LVLV MCKV MVLV PVKV PVLV PVNTV TVPV KVPV	E/AF/CS/AU E/SC E/AF/SC/AU/AM E/AF/SC E E/AF E E/AF/SC/AM E AU AF
<i>drink</i>	JVKV	E/SC/AU
<i>eat</i>	HVLV KVMV TVNV HVVV CVHV CVMV HVTV	E/SC/AU E/AF/SC/AU/AM E E/AF/SC/AU/AM E/AF/SC/AU/AM E/AF/SC/AU/AM E/A/SC/AM
<i>bite</i>	KVVV	E / SC / AM
<i>see</i>	HVCV WVTV KVRV NVKV	E/SC/AM E/AF/AM E/AF/SC E/SC/AM

	TVLV HVJV	SC/AU E
<i>hear</i>	HVLV CVMV KVLV	E/SC/AM E/AF/AM E/AF/SC
<i>know</i>	NVRV TVNV HNVN HVPV KVMV KVNTV KVKV KVPCV CVKV CVNV PVNV	E/SC/AU E/AU E/AF/SC E/AF/AU E/AF/SC/AU E/SC/AU E/AF/SC/AU/AM E/AF/SC/AU/AM E/AF/SC E/AF/SC AF/AU
<i>sleep</i>	HVMLV HVMV	SC/AM E/SC/AU
<i>die</i>	KVLV HVLV MVRV MVTV NVKV NVWV TVNV HNVN	E/SC E/AF/SC E/AF E/AF/SC E/AM E/AF/SC E/SF/SC E/AF/AM
<i>kill</i>	HVWV KVWV	E/AF/SC E/AF/SC
<i>swim</i>	-	-
<i>fly</i>	PVLV PVRV	E/AF/SC/AU E/AF/SC/AU
<i>walk</i>	CVHV HVMNV HVRV KVLV KVRV NVNV TVRKV WVTV KVJV HVJV	E/AF/SC E/AF/AM E/AF/SC/AU E/AF/SC/AU E/AF/SC/AU E/AF/SC/AU/AM E/AF E/AF/SC/AU/AM E/AF/SC/AM E/AF/AM
<i>come</i>	PVHV	AF/SC/AU/AM
<i>come</i>	PVHV	AF/SC/AU/AM
<i>lie</i>	KVLV KVWV	E/SC E/AF/SC/AU/AM
<i>sit</i>	TVNV	AF/SC/AU

<i>stand</i>	CVRV HVHV NVHV NVKV PVTV RVPV TVKV TVNV CVLV	E/AF/SC AF E/SC/AU/AM E/SC/AM AF SC E/SC/AM SC/AU E
<i>give</i>	TVHV CVHV	E/SC/AM E/AF/SC/AM
<i>say</i>	CVWV HVLV HVN HVWV JVNV KVRV KVWV MVHRV MVLV NVKV TVHV TVPV WVKV WVTV CVWV	E/AF SC/AU SC/AU SC/AF/SC/AM E/AF/SC/AM E/AF/SC E/AF/SC E/AF/AM AF/SC/AM E/AF/AM AF/SC/AM E/AF/SC/AM E/AF E/AF
<i>sun</i>	NVJV KVMV HVKV NVRV PVCV TVNV	E/AF E/SC SC/AM E/AU AF E/SF/AU
<i>moon</i>	TVLKV	E/AF/AU/AM
<i>star</i>	HVCRV TVCTV CVWV	E/AF/SC E/AF E
<i>water</i>	HVN KVHNV HVWV JVMV KVTV MVWV PVNV TVKV WVTV CVLV	SC/AM AF/AM E/AF/SC E/AF/AU/AM E/AM E/AF/SC/AU E/AF/AM SC/AU E/SC/AM E/SC

	HVKV	E/AF/SC/AM
<i>rain</i>	HVRCV	E/AF/SC/AM
<i>stone</i>	KVWV	E/AF/SC/AU/AM
	PVHV	E/SC/AU
	HVRV	SC/AU
	KVLV	E/AF/SC/AM
	LVNV	SC/AU
	PVNV	AF
	RVMLV	SC/AU
	TVHV	AU
	TVLV	E/AF/SC
	CVNV	AF/SC
	HVMCV	SC
	CVCV	SC
	HVNLV	AF/SC
	KVCV	E/AF/SC
<i>sand</i>	KVCV	E/AF/SC
<i>earth</i>	HVMGV	E/SC/AM
	KVRV	E/AF/SC
	PUMV	SC/AU
	TVHV	AF/SC/AU
	TVKV	E/AF/SC/AM
<i>cloud</i>	HVMKV	E/SC
	PVLV	E/AF
	NVPV	E/AF/SC/AU
<i>smoke</i>	CVNKV	E/AF/AU/AM
<i>fire</i>	HVMV	E/AF/AM
	TVLV	E/SC
	HVNKV	E/SC/AU/AM
	PVHV	E/AF/SC/AU/AM
	TVHV	E/AF/SC/AU/AM
	HVKV	AF/SC/AM
	CVCV	E
	HVHV	AF/AM
<i>ash(es)</i>	PVLV	E/SC/AM
	TVTV	E/SC/AM
	PVTV	E/AF/SC/AU/AM
<i>burn</i>	CVWV	E/AF/SC
	KVTV	E/AF/SC/AU/AM
	HVLTV	E/AF/AU
	HVRV	AF/SC/AU
	KVPV	E/AF
	KVRV	E/AF
	KVRV	E/AF
	LVKV	E/SC/AU/AM
	MVKV	E/SC/AU/AM

	PVKV PVLV PVLV PVRV TVKV WVRV KVN CVLV CVNV HVCV KVJV	E/AF/SC/AU/AM AF/SC/AM E/AF/AU/AM E/AF E/AF/SC/AM E/AF/SC E/AF E/AF/AM E/AF/AM E/AF/AM E/AF/SC/AU
<i>path</i>	HVRV LVMV RVMKV KVCV	E/AF/SC SC/AU SC/AU E/AF
<i>mountain</i>	KVN KVRV MVLV PVLV WURV PVRV TVKV TVLV TVPV	SC/AU E/AF/SC E/AF/SC AF/SC/AU E/AF E/AF E/AF/SC/AU/AM AF/AU E/AF/SC/AU
<i>red</i>	-	-
<i>green</i>	-	-
<i>yellow</i>	-	-
<i>white</i>	LVKV	E/AF/SC/AM
<i>black</i>	HVMV KVMV LVMV LVNV MVCV MVTV NVLV PVRKV CVLV CVMV CVNV HVPV KVRV	AF/SC/AU/AM E/AF/SC AU SC/AU E/SC/AM E/E/SC E/SC AF AF/SC E/AF AF/SC/AU E E/AF/SC/AM
<i>night</i>	LVLV HVDV NVNV KVCV	E/AF/SC E/AF E/AF/AM E/SC/AM

<i>hot</i>	KVCV LVMV TVPV	E/AF/AM E/AF/SC/AM E/AF/SC/AU
<i>cold</i>	CVRV HVKV KVMV PVCV KVRV	E/SC E/AF/SC E/SC E/SC E/SF/SC
<i>full</i>	KVPV MVHV MVLV PVNV TVKV TVM TVNKV TVRV HVPTV	E/AF/SC/AU AF E/AF AU/AM E/AF/SC E/AF/SC/AM E/AU E/SC E
<i>new</i>	CVRV HVRV MVHV NVWV WVLV	E/AF/SC E/AF/AU AF E/AF/SC E/AF
<i>good</i>	HVCV HVJV KVN NVKV WVNLV HVKV	E/AF/AM E/AF/SC E/SC/AM E/AF E/AF/SC/AM SC/AM
<i>round</i>	KVLV TVMPV	E/AF/SC E/AF/AM
<i>dry</i>	KVKV RVNKV KVRV CVCV KVCV KVLV KVMV	E/AF/SC SC/AU E/AF/SC E/AF E/AF/AM E/AF/SC/AM E
<i>name</i>	PVTV LVMNV MVKV CVMV	E/AF E/AF/SC AF/SC E/AF/SC

References

- Bickerton, D. (2014). Language Evolution: A Brief Guide for Linguists. *Lingua*, 117, 510-526.
- Bouchard-Côté, Al., Hall, D., Griffiths, Th. L., Klein, D. (2013). Automated Re-

construction of Ancient Languages using Probabilistic Models of Sound Change. In: *Proceedings of the National Academy of Sciences of the United States of America*, 110(11), 4224-4229. <https://www.stat.ubc.ca/~bouchard/pub/Bouchard2013Automated.pdf>.

Dubé, J. (2013). Reconsidering the “Isolating Protolanguage Hypothesis” in the Evolution of Morphology. In: *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society* (pp. 76-90). eLanguage.

Gibson, K.R. (2012). Language or protolanguage? A review of the ape language literature. In: M. Tallerman, K.R. Gibson (Eds). *Oxford Handbook of Language Evolution* (pp. 46-58). Oxford University Press.

Hurford, J. R. (1999). The Evolution of Language and Languages. In: *The Evolution of Culture* (pp. 173-193). University of Edinburgh Press. <http://www.lel.ed.ac.uk/~jim/dunbar.etal.html>.

*** (2014). *Indo-European Lexicon. PIE Etymon and IE Reflexes*. Linguistics Research Center, University of Texas. <https://lrc.la.utexas.edu/lex>.

Kiparsky, P. (1976). Historical Linguistics and the Origin of Language. *Annals of the New York Academy of Sciences*, 280, 97-103.

Starostin, G. (2014). Preliminary Lexicostatistics as a Basis for Language Classification: A New Approach. In: *Tower of Babel Project*. March 23. <https://starling.db.org/new100/Lexicostatistics.htm>.

Starostin, S. (2014). *The Tower of Babel. Homepage of Sergei Starostin*. <http://starling.rinet.ru>.

*** (2014). *Swadesh's 100-Word List*. Department of Linguistics. Portland State University (= S100WL, 2014).

The OEC: Facts about the language. Oxford Dictionaries. May 23, 2014. <http://elanguage.net/journals/bls/article/viewFile/3594/3491>.