*Svetlana Burlak*
Institute of Oriental Studies of the Russian Academy of Sciences, Moscow; svetlana.burlak@bk.ru

# Sources of lexical replacement and synonymy in Swadesh wordlists for proto-languages

In the practice of step-by-step reconstruction, situations are encountered where more than one word competes for the same slot in the protolanguage Swadesh-wordlist, and there is no way to dismiss one of the potential synonyms using standard methods. At the same time, in real languages it is almost always possible to choose which lexeme is the most suitable one for a particular slot in the wordlist. The difficulties in making such justifications for proto-languages may be related to both linguistic and sociolinguistic factors. Since we do not know the social landscape of the proto-language, as well as its evolution in descendant populations, there may not be enough material to draw conclusions about whether a word that was restricted to certain social situations in the proto-language (for example, used only by one gender, or only in relation to those higher or lower in the hierarchy, or only in conversations with children, etc.) could have become the primary term in one of the descendant branches after a change in the social structure. The practice of taboo leads to an increase in the number of synonyms in the language: at any given moment, one word is primary, but this can vary over time, so some descendant populations may inherit one word as primary, while others inherit a different one. Additionally, in descendant branches, phonetic or morphological alignments could have diverged in different directions, causing bases that once formed a single (possibly suppletive) paradigm to become the only representatives of the corresponding meaning in each of the descendant branches. The more time has passed since the separation of the proto-language, the harder it is to identify the alternations that caused the initial differences between the observable forms, so that they look apparently unrelated to each other. Perhaps the main difficulty in identifying the primary word lies in the fact that the proto-language may have had a differently structured nominative grid, so that contexts which appear homogeneous to the present observer were only considered so during the existence of the descendant languages, with some of them adopting one word as primary and others selecting a different one.

*Keywords*: Swadesh wordlist; synonymy; lexical replacement; comparative method; semantic reconstruction; language relationship; lexicostatistics.

In the practice of step-by-step reconstruction, situations are encountered where more than one word competes for the same slot in the Swadesh wordlist for the reconstructed protolanguage, and there is no way to dismiss one of the potential synonyms as derived, borrowed, areally restricted, or having undergone a more likely semantic shift. Nevertheless, in actually observed (currently existing and anciently documented) languages, it is almost always possible to justify the choice of a word for inclusion in the list by dismissing rarer or stylistically marked variants. In accordance with the principle of uniformitarianism, a similar situation should be assumed for proto-languages as well. Therefore, within the research conducted by the Moscow School of comparative linguistics, the task is set, first, to determine which of the potential synonyms was the primary one in the proto-language (Kassian et al. 2010), and second, to develop methods that should allow this to be done in all possible cases (see, for example, Starostin 2010: 100–110).

In fact, the only case of acceptable synonymy in a proto-list is so-called transitional synonymy, when "an older word is gradually being 'ushered out' by a more recent replacement" (Kassian et al. 2010: 48). In such a process, there is a period during which they are approximately equal in overall frequency (see Fig. 1), and different idiolects prefer different variants for the same contexts. However, at any given time, such cases are few (at most, 1–2 per 100-wordlist), so they do not significantly affect statistical calculations. Yet this situation highlights the following aspect of the innate human language capacity: humans are adapted to effortlessly acquire several different names for the same element of environmental *realia*. This adaptation was beneficial for small hunter-gatherer groups who, on one hand, communicated more frequently within the group than outside it (which facilitated language divergence through the mechanism of clustering human social networks, see Jackson 2019; Milroy & Gordon 2008), and on the other hand, were forced to communicate with representatives of neighboring groups for genetic exchange (the minimum viable population size for large primates is at least a thousand individuals; see, e.g., Frankham et al. 2014; Harcourt 2002), while the size of a human group sustaining itself by hunting and gathering is limited by environmental capacity and averages 25 people (Khrisanfova & Perevozchikov 2005: 108).

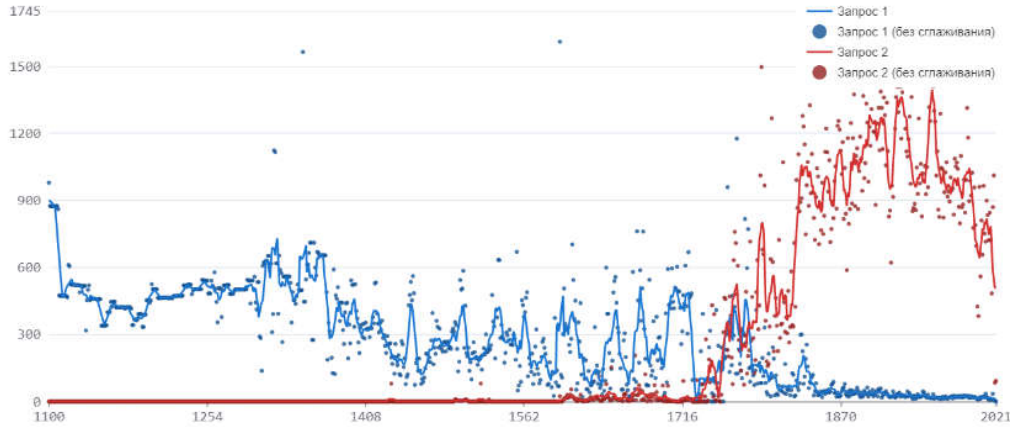


*Fig. 1.* Frequency of the lexemes *glaz* (red) and *oko* (blue) 'eye' in the Russian language of 12th–21st centuries AD (according to Russian National Corpus, see RNC). It can be seen that during the 18th century a lexical replacement takes place: the frequency of the word *glaz* grows rapidly while the frequency of the word *oko* decreases.

Nevertheless, even when collecting data from currently existing languages where direct elicitation from speakers is possible, many words on the list reveal insurmountable or poorly resolvable synonymy. In the case of a proto-language, where such opportunities do not exist, identifying the primary synonym can be challenging.

Difficulties in identifying the primary word filling the corresponding slot in the protolanguage wordlist may be related to both intra-linguistic and sociolinguistic reasons.

From a sociolinguistic perspective, any language can be represented by different social groups of speakers (by gender, age, occupation, etc.), as well as by different communication situations — and in all these cases, different names may be used for the same element of reality. For example, in modern Japanese, the idea of 'to give' is expressed by the verb 下さる *kudasaru*, if the giver is of higher status than the recipient, and by 差し上げる *sashiageru* (< originally 'to raise') in the opposite case. In Classical Ancient Chinese (see Starostin 2019: 159), the meaning 'to die' was expressed by the verb 薨 *hōng*, when referring to a high-ranking official, and by the verb 死 *sǐ* in other cases; the meaning 'to kill' was expressed by the verb 弑 *shì*, when referring to someone of higher hierarchical status (a prince, a father, etc.), and by the verb 殺 *shā* in other cases. Sometimes texts or recordings from different consultants allow one

to establish the most frequent and stylistically neutral variant (see, for example, the detailed philological analysis in Starostin 2019), but it is quite possible that compilers of dictionaries (especially older ones) may have overlooked such distinctions, and the inability to choose a single variant can affect reconstruction.

The influence of social hierarchy is particularly evident in personal pronouns: for example, in Japanese, different speakers in different situations use as the 1st person singular pronoun the words 私 *watashi* (in female and formal speech) / *watakushi* (in more formal speech), 僕 boku ((in male informal speech), 俺 ore (also in men's speech, in even more informal situations), etc. In Kassian et al. 2010: 65, 83, 85 it is recommended to exclude any polite forms when collecting pronouns — and this is feasible when collecting field data from a living language or during the philological analysis of an ancient language represented by an extensive text corpus. However, over the course of language development, polite forms can displace those that were originally neutral. For example, in standard Khmer, the word *khɲom* (etymologically 'slave') is a neutral 1Sg pronoun, while *ʔaɲ* (the original 1Sg word for 'I') is used as impolite or "as neutral in situations where expressing politeness is not required" (Pogibenko 2013: 30) (and therefore, accordingly, should not be included in the 100-wordlist).

The diversity of pronominal roots is not restricted to politeness forms. Thus, in the Kualan dialect of the Onya Darat Austronesian language (Borneo), spoken by relatively small communities, the choice of pronoun depends on how the generation of the speaker and the generation of the listener relate: for example, when speaking to one's child (or nephew), a person calls themselves *maaq*, and their interlocutor *omo*; if a person is speaking to their brothers, sisters, or parents (as well as their brothers or sisters), they call themselves *oko*, and the interlocutor *omo* (if of the same generation) or *okam* (if of an older generation, and it is the generation, not age, that matters; see Tadmor 2015: 84–85).

Even if some of these variants are accepted as the most neutral in some synchronic state of the language, it is quite possible that with a change in lifestyle, everything can change: for example, when speakers of this dialect move to cities where they have to communicate with persons whose generation they do not know, they develop some system that is different from the original one (Tadmor 2015: 96) — and in different places, such changes can proceed in different ways. Changes in lifestyle are also possible in non-urban cultures living by hunting and gathering or primitive agriculture (see, e.g., Henrich 2016: 174–175).

In the languages of Southeast Asia, pronouns are used less frequently in principle — if a participant of a communication act can be designated by kinship, seniority, title, clan affiliation, etc., the corresponding noun will be used in the sentence (Tadmor 2015: 79). Due to such practices the probability of borrowing personal pronouns increases (which would be undetectable if the source language becomes extinct), and so does the probability that some pronouns may be replaced by nouns.

Another sociolinguistic reason for synonymy is the difference between male and female speech. For example, in Japanese (Holmes & Wilson 2022: 224), women's main word for 'to eat' is 食べる *taberu*, while men's word is 喰う *kuu*; the meaning 'stomach' is expressed as お腹 *onaka* by women and as 腹 *hara* by men (although the word 'stomach' does not belong to the 100-wordlist, frequent patterns of semantic shift imply that it can always become the main designation for 'belly'). In the Kokama-Kokamilla language (Tupi family, South America), men would say ***r**a* '(s)he', ***r**aepe* 'then', ***r**amua* 'other', while women would use words ***y**a* '(s)he', ***y**aepe* 'then', ***y**amua* 'other' (Vallejos-Yopán 2015).

Male and female variants may often differ phonetically. For example, in Bengali, in some words, women pronounce *l*, while men pronounce *n* (Holmes & Wilson 2022: 224); in Chukchi, male variants with *r*, *č*, and *rk*, *rg*, would until recently be respectively pronounced by women

as *c* and *cc*, e.g., the word 'walrus' in male speech was ***rərkə***, and in female speech ***cəccə***, etc. (Skorik 1961: 33); this state of affairs may be caused by men often marrying representatives of a neighboring dialect or a closely related language (Dunn 1999: 27 with lit.).

In a number of cases, the male and female variants differ morphologically. For example, in the Yana language (N. America), men add suffixes, so the meaning 'person' is expressed by the form *yaa* in female speech and the form *yaana* in male speech; the meaning 'fire' is expressed by the forms *ʔau* and *ʔauna* respectively (Holmes & Wilson 2022: 224). In the Yanyuwa language (Australia; Holmes & Wilson 2022: 224), men use different class prefixes than women, e.g., for 'meat' women use the form *ni-warnnyi*, and men *na-warnnyi*. In the Lakota language (White Hat 1999: 18), men and women use different enclitics, e.g., the meaning 'this is good' in female speech is rendered as *wašte kšto*, and in male speech as *wašte yelo*.

In male and female speech, the same language elements can occur with different frequencies: for example, in Canadian French, both women and men sometimes drop *l* in the constructions *il y a* and *il fait* (as in *il fait beau* 'the weather is good', *il fait froid* 'it is cold' etc.), but men do it more often (Holmes & Wilson 2022: 227).

It is clear that in a developed literary language with a large text corpus, careful philological analysis will allow the most neutral variant to be identified; but if such a practice exists in a language used exclusively in everyday communication situations (and there is no reason why such a language could not later become a proto-language for some group), choosing a more neutral variant from male and female or higher/lower one may be difficult (moreover, when working with a limited number of consultants, only one of the existing variants may be included in the dictionary).

Synonyms can originate from so-called "child-directed speech" (or "nursery laguage", which is "a language subsystem considered convenient to talk to small children" (Yeliseyeva et al. 2017: 7). Words from this register can displace the original words, surpassing them in frequency, cf., for example, the frequencies of the words *mama* and *mat'* 'mother' in modern Russian (according to the Russian National Corpus).

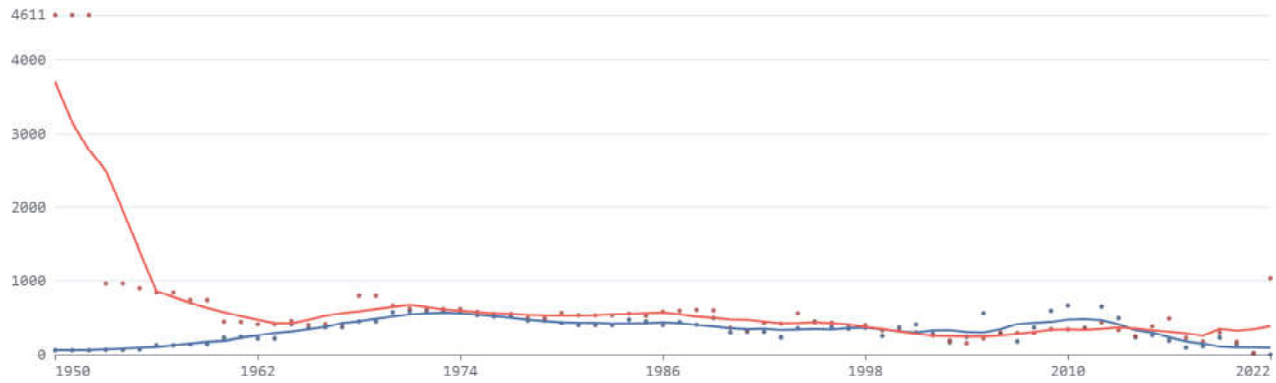


*Fig. 2.* The frequency of the lexemes *mama* (blue) and *mat'* (red) 'mother' in the Russian language in 20th–21st centuries (according to Russian National Corpus).

It can be seen that at certain time intervals the word *mama* is used more frequently than *mat'*. (Note that, although the word 'mother' does not belong to the Swadesh 100-wordlist, it can serve as a source for the word meaning 'woman', cf. Lith. *móteris* 'woman' < PIE **méh$_2$tēr* 'mother'). Likewise, the word for 'breast' also can originate from the child-directed speech (Starostin 2010: 93).

Nursery forms are not obliged to be phonetically similar to the corresponding forms in standard adult language (cf., e.g. Rus. *mashina* and *bibika* 'car', *jest'* and *am-am* 'to eat', Engl.

*tummy* ‘belly’). Since such subsystems exist in many languages, it is probable that they existed also in proto-language time and could supply the sources for lexical replacement.

If a culture has a developed practice of tabooing, for many meanings there can be sets of complete synonyms, allowing one to freely replace one with another if necessary. For example, in the Misima language (Misima Island, Papua New Guinea), where the names of all deceased, as well as words similar to them, are tabooed, there are five words meaning ‘fire’ (Simons 1982: 203). At the same time, different lexemes can act as the most frequent and neutral words at different times (and for different speakers). P. Black notes (Black 1997: 57) that many lexemes of various Australian languages, which in the 1930s had been cited as examples of tabooed words, were recorded in dictionaries of the 1980s as being in active use. On Tahiti island at the end of the 17th century, the word *pō* ‘night’ (< Proto-East-Polynesian **pō*) became tabooed because it was part of the dynastic name of the island’s rulers, Pomare, and the word *ru‘i* (originally ‘dark, blind’) began to be used instead, but in the 19th century, some speakers switched back to use *pō*, while others continued to use *ru‘i* (Belikov 2006: 24–25).

Theoretically, there is a possibility that the Proto-Indo-European lexemes **péh₂wr̥* and **h₁n̥gʷnís* ‘fire’ could have been such synonyms, but then in some descendant languages one of these words became the main designation for fire, while the other one chose a different option (thus, if only the Italic and Indo-Iranian branches of this family had survived, the lexeme **h₁n̥gʷnís* would be recognized as the main Indo-European word for ‘fire’, since it exists in both branches). Note, however, that such hypotheses should not be accepted without considerable proof (see below).

Furthermore, different speakers may treat different synonyms as stylistically neutral variants, and when a population splits into two, the percentage of those who consider one or another variant stylistically neutral may be unequal in different parts; as a result, one daughter population most often will choose one synonym as stylistically neutral, while the other one will prefer the alternative. During reconstruction, it will then be impossible to determine which variant was more often chosen as stylistically neutral in the original population (and which, accordingly, should be considered the most stylistically neutral for the proto-language).

A word that was the most common at some point may over time move to the linguistic periphery and then return to the status of the main carrier of the corresponding semantics. Perhaps such was the situation with the Indo-European root **gʰel-*, expressing the meaning ‘yellow’ (> Skt. *hári-* ‘yellow, green’, Greek χλωρός ‘green, yellow, greenish-yellow’, Rus. *жёлтый* ‘yellow’). In Latin, a replacement occurred: instead of this root, the main designation for ‘yellow’ became the word *fulvus* (GLD 2025), while the derivate of the root **gʰel-* — *galbinus* ‘greenish-yellow’ — moved to the periphery; however, it is precisely this word (and not *fulvus*) which is inherited as the main designation of yellow color in the modern Romance languages (French *jaune*, Romanian *galben*), cf. Normanskaya 2005.

If a community has widespread diglossia (this may also happen in non-literate cultures) and one of the languages is used as the main means for everyday communication, it lacks a high register, and accordingly, words of the lower register become the most frequent. Such processes may trigger for certain words a shift to ‘stylistically neutral’ from (possibly) initially ‘pejorative’, cf. Tok Pisin *klok* ‘heart’ (Belikov 2006: 94) (< Engl. *clock*), Proto-Slavic **rǭkà* ‘hand’ (from a verb meaning ‘to gather’, cf. Lithuanian *riñkti*), Khmer (Surin dialect) *ŋɯəp* ‘to die’ (cf. standard Khmer *ŋiəp* ‘to wither (of a plant), perish (of an animal), die (of a person, pejorative)’ (Pogibenko 2013: 30). If the idiom occupying the position of a higher register is closely related to the one used in everyday communication, synonymous lexemes can appear in the same contexts for the same speakers. Thus, in examples given in the dictionary of the Deulino village dialect of Russian, the lexemes *зенçина* ‘woman’ and *baba* ‘woman (colloquial/pejorative)’

appear in the same contexts, cf. *dodʲelʲistaja ʒençina* and *dodʲelʲistaja baba* ‘shrewd woman’ from the same consultant within the same text (Ossovetskiy 1969: 145). When compiling the Swadesh wordlist for this dialect, the variant *ʒençina* was chosen (GLD 2025), but there is a chance that a different selection of consultants for the dictionary (or different research attitudes of the list compiler) would lead to a different result.

In general, dialects are characterized by a situation where features of one’s own and neighboring idioms are distributed in speech with some non-zero frequency, which can increase and decrease for various reasons. The following situation, documented in one of the Vladimir-Volga dialects, is indicative: “In 2012, immediately after meeting, the first conversation with EM (over two hours) was recorded. The proportion of examples with *yokan’ye* (a phonological feature — S.B.) in this recording is small: 22% <…>, and the number of examples with *yokan’ye* noticeably increased towards the end of the conversation (apparently as EM’s communication with dialectologists became more relaxed). In this conversation, EM approached the first group of consultants in terms of *yokan’ye* — persons with non-peasant professions. In 2014, we visited EM twice; both times she received us very hospitably, as old acquaintances. Most of the first conversation was casual talk, and towards the end, EM was offered a questionnaire on the declension and stress of masculine nouns. In this conversation, the proportion of forms with *yokan’ye* turned out to be significant: more than 55%, i.e., “at the level” of women of her age with incomplete secondary education, farm workers (group 3). The next morning, the questionnaire survey was continued, but EM was already focused on problems of speech correctness and throughout the recording discussed with us the difference between dialect forms and standard ones. For her, this was a conversation on professional topics with fellow philologists. The percentage of forms with *yokan’ye* in this conversation turned out to be lower: 33% (42% in place of **e*), i.e., between the data for the first and second groups of informants. Thus, EM partly controls *yokan’ye* in her speech: it becomes less prominent in a «semi-official» communication situation (with unfamiliar visitors from Moscow) and in a «professional» communication situation. The proportion of examples with *yokan’ye* in EM during more relaxed communication corresponds to the archaic variety of dialect usage. Possibly, other consultants with a low proportion of *yokan’ye* are capable, like EM, of controlling its use, however, we do not have data on this" (Dyachenko et al. 2018: 56).

If the language has a developed folklore tradition, it can preserve more archaic lexemes. For example, Tuamotu preserves an archaic style called *fangu* (the language of sacred chants), where many words that have been replaced in ordinary speech due to taboo are preserved: thus, the meanings ‘water’ and ‘to drink’ in ordinary Tuamotu are expressed by the word *komo*, while in *fangu* the original words *vai* ‘water’ and *inu* ‘to drink’ are preserved; for ‘big’, not the standard *tooreu* is used, but rather the original *nui*; for ‘night’, not the standard *ruki*, but the original *poo*, etc. (White 1968: 65). When a population splits, the stylistic attitudes of speakers can change — differently in different parts.

The listed situations are relevant for the case when a proto-language has split into two daughter branches, and one variant is chosen in one of them, and the other in the other one. But, generally speaking, such “semantic criss-crossing” (using the terminology of the Moscow School of comparative linguistics, see Starostin 2010: 103–104) can also be seen in cases when in each of the groups both variants are present. This may be a consequence of the law of homological series in hereditary variability, originally formulated by N. I. Vavilov for plants, but apparently relevant for any objects evolving naturally. This law states that “species and genera that are genetically close are characterized by similar series of hereditary variability with such regularity that, knowing the series of forms within one species, one can predict the finding of parallel forms in other species and genera” (Vavilov 1935).

If the language experienced several periods of dialect fragmentation and subsequent "koineization", traces of such stylistic differences, as well as cases of transitional synonymy at different stages of its existence, may, after a significant time, look like proto-language synonymy. The matter can also be complicated by the fact that when borrowing from closely related idioms, phoneme-by-phoneme recalculation is often applied (Burlak & Starostin 2005: 63), which makes borrowed words indistinguishable from inherited ones and masks lexical replacements.

Synonyms for concepts frequently encountered in everyday life can sometimes be created literally out of nowhere. As was already noted by Igor Dyakonov (1984: 8), concepts perceived as particularly important for a particular culture, the ones that are often talked about, usually have many synonyms. One can give an example from modern Russian: cat lovers, who often have to talk about their pets (and who usually experience strong positive emotions towards them), call them not only by the common words *koshka* 'cat' and *kot* 'tomcat', but also by such words as *kosha*, *kotyara*, *kotik*, *kotejka*, *kotofej*, *kote*, *kotenitsa*, *kotyandra*, *koshundra*, *koshandra*, *koshendra*, *koshak*, *koshuk*, *koshakan*, *koshkatan*, *kosharik*, *koshara*, *koshanya*, *koshe*, *koshko*, *kotanych*, *kyso* (Iomdin 2023).

Such an abundance of variants apparently has a neurobiological basis: research shows that, if a word is repeated too often, it loses its meaning — both for the speaker and the listener: as experiments show, after about 30 repetitions of the same word in a row, it suddenly begins to seem meaningless (the "jamais vu" effect, as described in Moulin et al. 2021). This is not surprising: extinction of response to a too frequently repeated stimulus is characteristic of all living organisms (starting with unicellular ones). When producing and perceiving rare words, additional neural circuits are involved (Black 2003; Vlasova et al. 2015; Malyutina et al. 2012).

Moreover, for different speakers, different words will represent the main (most frequent and stylistically neutral) variant, and at times this can vary even for the exact same speaker. Whether such an abundance of names extends only to cultural vocabulary or can also affect some basic meanings is unknown.

In addition to such cases of true synonymy, where words truly have the same meaning, and frequency and stylistic characteristics differ either among different speakers or in different communication situations, there are a number of situations where either the same original root becomes unrecognizable due to phonetic or morphological changes, or two different roots, initially differing in meaning or morphological properties, lose these differences, and different descendant languages choose either one or the other as the main one. Let us consider these cases in more detail.

As a result of phonetic changes, a root can lose some phonemes, and then reanalysis can occur. For example, in modern Spanish, the meaning 'to eat' is expressed by the root *com-*, cf. *comer* 'to eat', *comida* 'food' (this same root is present also in Portuguese and Galician). This happened because the Latin root *ed-* 'to eat' (inf. *edere*) underwent the loss of *d* in most forms, and at some point speakers, to ensure better understanding, began to replace *edere* with *com-edere* 'to eat up'. The root *-e-* in *comer* ceased to be distinguished, and the former prefix *com-* ('with') became the carrier of the meaning 'to eat'. If only Spanish, Portuguese, Galician, and, for example, Slavic or Germanic languages remained from the Indo-European family, the researcher would have to deal with proto-language synonymy: two roots would be reconstructed for the meaning 'to eat', **ed-* and **com-*, with no observable semantic difference between them.

Roots are especially affected in languages which, due to a restructuring of phonotactics, lose polysyllabicity. For example, in Mon-Khmer and Kra-Dai languages transitioning from polysyllabism to monosyllabism, initial syllables ("presyllables") become "weak": they cannot

appear alone, without a strong syllable nearby, and are characterized by a number of restrictions — they have no tone, no consonant clusters are possible at the beginning, the vowel is reduced, and there is no final consonant. This leads to a large number of phonetic changes in initial syllables.

Furthermore, in such presyllables, variability like "*khr- ~ khl-*, *phr- ~ phl-*, *khr- ~ phr-*, etc." (Samarina 2007: 46) is often noted, with different speakers possibly preferring certain variants or using both freely (ibid.). For example, in the Ruc language (Vietic group of the Mon-Khmer branch of the Austroasiatic family, Vietnam) in words like *rata*$^{3}$ ~ *lata*$^{3}$ 'stone' or *rəka*$^{1}$ ~ *təka*$^{1}$ 'chicken', free variation of the initial consonant is observed (ibid.: 43). If such variation occurred at the moment of the proto-language's binary split, one variant could gain greater frequency in one daughter branch, and another one in the second one.

Such presyllables could be former prefixes, and in this case, despite the variability, roots could be reconstructed correctly, allowing cognates to be identified. But if root morphemes stood in the initial position in the past (and in the final position there were either also root or suffixal morphemes), then the more time has passed between the split of the proto-language and its descendants available for study, the less likely it is to determine the original identity of the roots, and the higher the probability that reflexes of the same proto-language root will look like etymologically unrelated synonymous morphemes without any semantic distribution.

The situation would look the same if the original phonetic differences were associated with male / female speech, politeness, taboo, or were stylistic variants. Furthermore, the more time has passed since the split of the proto-language, the more phonetic changes have occurred and the more lexemes have been lost, so the initial conditions of phonetic changes may be impossible to establish due to the material being too limited.

False synonymy can be due to morphological reasons. For example, in Atlantic languages (a West Aftrican family), the noun class marker can be prefixal, suffixal, or circumfixal. Thus, in the Bapen language (Tenda group of the northern branch), the word for 'tree' is *a-tu-ã̀* (Pl. *ɓa-tu-ɓã́* indicates that its root is *tu*); in other languages of the Tenda group, class markers are represented by prefixes, cf. Basari *a-tʌ́x* (Pl. *ɓa-tʌ́x*), Bedik *ga-tɔ́* (Pl. *ɓa-tɔ́*) 'tree' (Pozdnyakov 1993: 52). Markers represented as suffixes in Fula (Fula-Serer group of the northern branch) appear as circumfixes in Serer (ibid.: 27–31). This state of affairs may be a consequence of grammaticalization processes. For example, in Swahili, the expression 'these Swahili people' looks like *wa-suahili ha-wa*, where the class marker *wa* appears both at the beginning and at the end of the noun phrase. With the development of grammaticalization according to the standard model, first the demonstrative pronoun turns into a definite article, then the article begins to accompany every noun and becomes simply a noun marker (see, e.g., Kuteva et al. 2019: 27). In doing so, the corresponding morpheme undergoes phonetic and prosodic erosion. Subsequently, redundancy can be reduced in one way or another: either the initial or the final part of the marker can be omitted. Taking into account the fact that the class marker can be determined by the phonetic characteristics of the lexeme (Sumbatova 2022), a situation of reanalysis is quite realistic. Thus, for the word 'tongue' in Proto-Atlantic, the form **d-em-əd* ~ **r-em-ər* is reconstructed (Pozdnyakov 1993: 146–147); subsequently it developed into **de-məd* (root **məd*) in some languages and into *dem-əd* (root **dem*) in others, cf. Kisi *demu-le* (Pl. *demu-ã̀*), Landuma *da-mer* (Pl. *sə-mer*). The more time has passed since the split of the proto-language, the less material researchers have to trace the grammaticalization process, and accordingly, the more likely it is to obtain two seemingly unrelated synonyms like **lem* and **mel* during reconstruction.

Word formation can also be a source of false synonymy. For example, in Nicobarese languages, the concept 'palm (of hand)' is conveyed by the words *el-ti:* (Car language) and *mòh-ti:* (Teressa language). The second part in these lexical complexes goes back to the same proto-

language root, while the initial parts are different: *el* means ‘inner part’, *mòh* means ‘tip, protrusion, nose’ (Pogibenko 2018: 203). Such formations often represent, in essence, so-called “micro-syntactic” constructions where components can be freely replaced by similar ones. The more time has passed since the split of the proto-language, the higher is the probability that in at least one of the daughter branches, the part that represented the proto-language root was lost — for purely phonetic reasons — and then in the reconstruction, seemingly unrelated lexemes like *elt* and *moh* will appear. Languages prone to this type of word formation often repeat the corresponding processes as the semantic motivation of the previous lexical complex becomes obscured. For example, in modern Khmer, the word *crʊmuh*, which arose from the lexical complex ‘hole’ + ‘nose’ (< Old Khmer *muh* ‘nose’) and originally meant ‘nostril’, came to mean ‘nose’, and for the meaning ‘nostril’, a new lexical complex of words with the same meaning is used: *rɔnthɛə? crʊmuh*, where the second component means ‘nose’, while the first goes back to a borrowing from Sanskrit *randhra* ‘hole’ (Pogibenko 2018: 203–204).

Perhaps cases of this type, like the case of Spanish *com-* ‘to eat’, should be treated as lexical replacements. But it is not entirely clear whether replacements of this type have the same speed as replacements occurring due to semantic shifts, or a different one.

Another reason for the appearance of synonymy in a proto-language wordlist can be suppletion. In a proto-language — just as in any extant language — words could have a suppletive paradigm, cf. English *go* and *went*, Tocharian A *y-* (Pres.) / *kälk-* (Pt.) ‘to go’. The paper Kassian et al. 2010 proposes to choose the morphologically simplest stem. Sometimes this approach may conflict with the principle of choosing the most frequent variant, since for different lexemes, various morphological forms can have unequal frequency. So, for example, according to data from the SynTagRus corpus (part of the Russian National Corpus, see, e.g., Timoshenko et al. 2021, for the word *vjetka* ‘branch’ the most frequent form is the instrumental plural, for the words *gora* ‘mountain’ and *ruka* ‘hand’, the prepositional plural.

In accordance with the principle of choosing the most frequent of the synonymous designations for the same concept, “for telic verbs, apparently, the perfective aspect form should be taken, and for atelic verbs — the imperfective aspect form: the frequencies of the corresponding dictionary units differ very much” (Burlak 2021: 298).

At the same time, in different sets of texts (and, apparently, for different consultants), frequencies can have even opposite distributions. For example, in modern standard Russian, according to data from the Russian National Corpus (as of February 16, 2025), the forms *chelovek* ‘person (sg.)’ and *lyudi* ‘people (pl.)’ (in all cases together) have approximately equal frequency: *chelovek* occurs 553,208 times, *lyudi* — 571,061 times, but in the Old Russian corpus, the plural *lyudie* (1445 examples) occurs twice as often as the singular *chelověkъ* (713 examples), while in the Dialect corpus the ratio is reversed: for 410 uses of the singular *chelovek*, only 7 examples of the plural *lyudi* were found. When a population splits (which is necessary for the split of a proto-language), speakers who more often use one of the suppletive morphological forms could predominate in one of the daughter populations while people of the other one could more often use the other form; subsequently, this could lead to paradigm unification (in one daughter population based on one stem and in the other, based on its counterpart), and in this case, during step-by-step reconstruction, two seemingly unrelated synonyms will compete for the same slot in the proto-language wordlist. The more time has passed since the split of the proto-language, the less chance there is to find any evidence that in the proto-language both corresponding stems were present and formed one suppletive paradigm with such-and-such distribution of forms by grammatical meaning.

The grammatical meanings by which suppletive stems are distributed can be very different. For example, in many languages spoken on the northwestern coast of North America, stems

used for singular and plural subjects in the verb are suppletive (Campbell 2004: 304–305). The verb ‘to give’ can distinguish stems depending on who is given to.

Initially, suppletion could have developed due to the rarity of phonetic alternation. For example, for Old Church Slavonic, the forms of the infinitive *gъnati* ‘to drive, to chase’ and the present tense *ženetъ* can be meaningfully considered as suppletive (in Polivanova 2013: 72, 430, 588 they are recognized as allomorphs with an anomalous relation), despite the fact that they exhibit regular ablaut: *e*-grade in the present, zero-grade in the infinitive. The more time has passed since the split of the proto-language, the less likely it is to find enough material to discover the rule for the distribution of allomorphs.

The most frequent source of false synonymy is semantics: words can be replaced due to connotations (like the Latin designation for ‘black’, *ater*, was replaced in all Romance languages, see Normanskaya 2005: 180) or due to emphasis: words that had emphatic highlighting could lose it, becoming the most frequent and neutral designations of the corresponding concepts (in this case, new emphatically highlighted words could take their place), or be lost and replaced by new emphatically highlighted words. In different daughter branches, these processes could proceed differently, and the more time has passed since the split of the proto-language, the less chance there would be to trace this, and then two synonyms will compete with equal grounds for a slot in the proto-language wordlist: both the one that was originally primary and the one that was originally stylistically marked.

Furthermore, any language segments reality into fragments denoted by individual words in ways that are distinct from any other language. In the paper Kassian et al. 2010, trying to formalize the principles of compiling 110-item wordlists and develop a meticulous algorithm for identifying the most basic meanings from a set of close ones, contexts are provided that should help the researcher choose the most basic lexeme from those with similar meanings. For example, for the meaning ‘to lie’, it is proposed to take the “verb of continuative/durative action. It is generally not recommended to fill in the slot with the inchoative verb ‘to lie down’, although such practice is rather frequent (and, in truth, both notions are quite commonly expressed with the same root). Should be strictly separated from lying in particular postures (‘lie on one’s side’, ‘lie face down’ etc.)” (Kassian et al. 2010: 66). For modern languages, where consultant questioning is possible, making such distinctions is quite feasible, but in the process of language development, any one such verb (‘to lie down’, ‘to lie on one’s side’, ‘to lie face down’, etc.) could undergo a semantic shift in one of the daughter branches and become the main verb for the meaning ‘to lie’, so that if the corresponding root is lost in other branch(es), the semantic development between the descendant of the proto-language word ‘to lie’ and the proto-language word ‘to lie down’ (or similar) may be impossible to establish.

Words for ‘man’ and ‘woman’ can originate from kinship terms (see above), and kinship terms in many languages are subdivided into referential and appellative. For example, in Tamil, the meaning ‘mother’ is expressed with “the two main roots, *ammā* and *tāy*, of which the root *tāy* is referential, and *ammā* is appellative” (Smirnitskaya 2024: 64). If a long time has passed since the split of the proto-language, semantic shifts could have occurred in the daughter populations, as a result of which the meaning ‘woman’ originated in one language from that word for ‘mother’ which in the proto-language was appellative, and in another — from the one that was referential; in this case, during reconstruction, two absolutely equal synonyms will compete for the slot ‘woman’, and even if it would be possible to establish that each of them came from a word meaning ‘mother’, it would not be possible to choose the “most basic” lexeme.

Contexts formulated on the base of the study of a large number of languages allow many semantic distinctions to be removed, but hardly all of them. For example, for the meaning

'tail', the authors of Kassian et al. 2010 propose to take the word that applies to mammals (thereby dismissing separate terms for the tail of a bird, fish, etc.) and is used in contexts: "1. Animals have tails, man does not have a tail. 2. (The beast) is swishing its tail to drive away the flies. 3. He cut off (the beast's) tail" (Kassian et al. 2010: 79). However, a language may have an opposition between a fluffy tail (like a fox's) and a non-fluffy one (like a cow's) or similar. For example, in Icelandic, mammals have tails such as *dindill* 'short tail (like a sheep's or seal's)', *tagl* 'horse's tail', *hali* 'tail with a tuft at the end (like a lion's or cow's)', *rófa* — one of the most common tails, such as a cat's, mouse's, or fox's (but its first meaning, according to the dictionary, is not 'tail' but 'coccyx') and *skott*, the same as *rófa*. In the context "to cut off the tail", any of these is acceptable. If we choose the most frequent variant, then one must choose *rófa* (if its first meaning would not prevent this) or *skott* (if focusing more on semantics than frequency), and if we rely on the context "wave the tail, driving away insects", then *hali* should be preferred (as in the database GLD 2025) or *tagl*.

It is not always helpful to use context in order to remove the distinction between a large (usually predatory) bird and a small bird: the contexts "1. Something is moving in the bushes, I cannot tell if it is a bird or an animal. 2. Birds lay eggs, animals and people bear children. 3. There is a bird flying on high, I cannot tell what kind", recommended in Kassian et al. 2010: 52, may not yield a result, since distinguishing a far-soaring eagle from a closely fluttering sparrow is not difficult for people who know nature well. The context "Birds lay eggs, animals and humans bear children" may appear confusing, just like Russian or English speakers would find confusing a context in which a researcher accustomed to grouping birds with reptiles would ask them to translate how they (diapsids) all lay eggs.

The 400-item wordlist of basic vocabulary, currently undergoing testing as part of the Moscow School's "Comparative Onomasiological Database for the Ancestral States of Eurasian Linguistic Families" (https://starlingdb.org/new400/), includes other meanings that in languages can be represented by more than one word. For example, the meaning 'to fall' (also included in the "Leipzig-Jakarta wordlist" as proposed in Haspelmath & Tadmor 2009) can be expressed in four different ways: "movement from a higher surface to a lower one (*the vase fell off the table*), loss of vertical orientation (*the vase fell and the water spilled on the tablecloth*), falling-collapse (*the house fell down*) and detachment (*the dress fell off the hanger*)" (Reznikova et al. 2020: 9). The meaning 'to love (someone, not something)' in Ancient Greek is conveyed by 4 verbs (initially, apparently, corresponding to relationships in the nuclear family), cf. στέργω (initially describing love of parents for their children), ἀγαπάω (initially describing love of children for their parents), φῐλέω (initially describing love of siblings for each other) and ἐρᾰ́ω (initially describing love of the spouses for each other).

In a number of cases, the possibility of the absence of a single basic lexeme with the necessary meaning is recognized in Kassian et al. 2010. For 8 words from the 110-word list, the paper recommends to accept synonyms if there is no general term, namely:

1) 'to fly', if there are different roots for an iterative verb 'to fly' (*Where is the bird flying?*) and a habitual verb 'to fly' (*Little chicks do not yet know how to fly*);

2) 'to give', if a language makes a clear and regular distinction between 'to give (to 1st/2nd person)' and 'to give (to 3rd person)';

3–4) 'man' and 'woman', if a language has different terms for members of different age groups (in such cases, the respective lexemes for the age of ca. 20–45 are allowed to be treated as synonymous);

5) 'that', if a language has a more elaborated deixis system;

6) 'not', if different tense / aspect etc. forms use different negative markers and no preferences can be easily formulated;

7) ‘mouth’, if a language possesses two separate terms, one for ‘oral cavity’ (*His mouth is full of water*) and another for ‘external parts of the mouth’ (*He has a very wide mouth*);

8) ‘tooth’, if a language has separate terms for ‘back tooth, molar’ and ‘front tooth, incisor’.

Other 5 slots are supposed to be split into several: these are pronouns (‘I’ and ‘you’, where differences between direct and oblique stems are common, and ‘we’, where besides this difference, a clusivity distinction is often also present), the words ‘round’ (where characteristics of disc-shaped and spherical objects are often distinguished) and ‘thin’ (where terms characterizing flat and elongated objects may be distinguished, like a thin sheet and a thin rope). For the word ‘warm’, the possibility of splitting into ‘warm’ and ‘hot’ is provided, and for the word ‘year’, the possibility of not filling the slot in the absence of a general term is envisaged.

The grounds for distinguishing some *realia* from others, naming them with separate words, can be very diverse, and if the intuition of the list compilers and native speakers does not coincide (or semantic distinctions are not well researched), words with differing meanings can look like synonyms. For example, Russian has the words *pepel* and *zola*, both meaning ‘ashes’. In the contexts proposed in Kassian et al. 2010: 51 (“1. The campfire has left only ashes. 2. The wind scattered the ashes. 3. He scooped up a handful of ashes”), both are acceptable, with different speakers producing different variants as their first reaction, and neither of them in these contexts appears to be more frequent than the other. Cf. (about a person burning money in a fireplace): “Bumping into something hot, he would pull his hand back — and again thrust it into the fireplace. Having gathered a smoldering pile of ashes (*zola*), he carried it through the bedroom to the bathroom, carelessly spilling it on the floor... In two goes, he transferred almost all the ashes (*pepel*) without residue in the fold of his robe and with it many small embers, from which the robe smoked” (M. A. Osorgin “Svidetel’ istorii” 1932, cit. via RNC. Translated by DeepSeek — S.B.); “A slow rain of ashes (*zola*) fell. It covered the river water with a gray coating. Sometimes birch leaves, turned into ashes (*pepel*), flew down from the sky” (Paustovsky, “Zhelty svet”, 1936, cit. via RNC, translated by DeepSeek — S.B.).

The greater frequency of the word *pepel* (it is used 4847 times in the main RNC corpus, while *zola* is only encountered 2893 times) is likely provided by collocations like *vulkanicheskiy pepel* ‘volcanic ash’, *prevratit’ v pepel* ‘to turn to ashes’, or *posypat’ golovu peplom* ‘to pour ashes on one’s head (in mourning)’. In the Old Russian corpus there are 26 examples of *pepelъ* and its variant *popelъ*, while *zola* does not occur at all (although the lexeme *zola* is of Proto-Slavic origin); in Old Russian there are 50 examples of *pepelъ* and 49 examples of *zola*.

From a fire or conflagration, both *pepel* and *zola* remain in roughly equal measure. Both could be used for treatment, cf. “Take a live lizard and burn it to ashes (*pepel*). Then sprinkle those ashes (*pepel*) on the horse’s sore…” and “And however many years old the horse will be, take that many fresh eggs and burn them and make ashes (*zola*). And mix it with that milk and pour it into the horse’s mouth, it will be healthy” — both examples are from a medical book on horses dating to the third quarter of the 17th century, cit. via RNC (translated by DeepSeek — S.B.). With the spread of tobacco smoking, the collocations *tabachnyj pepel* and *tabachnaja zola*, both meaning ‘tobacco ash’, initially competed in roughly equal measure: in the RNC from the early 19th century to the present, there are 28 examples of “tobacco *pepel*” and 20 of “tobacco *zola*”), but in the 20th century, *zola* loses ground: *papirosnaja zola* ‘cigarette ash’ ends in 1938, and *tabachnaja zola* ‘tobacco ash’ — in 1980.

Nevertheless, careful analysis of contexts shows that these lexemes are not fully synonymous. Cf.: “And what’s the difference: *pepel* and *zola*? Which is cleaner? Which is more terminal? *Pepel*, of course — *zola* is even used for fertilizing” (M. I. Tsvetaeva “Istorija odnogo posv’ashchenija”, 1931, cit. via RNC, translated by DeepSeek — S.B.). According to RNC data, only *pepel* can be volcanic, and only *zola* can be taken from a stove. Only *pepel* can fall off

(*osypat'sya*), fall (*padat'*), and fly (*letat'*), but only *zola* can be produced (*poluchat'sya*), serve (*sluzhit'*), and contain smth. (*soderzhat'*); only *pepel* can be shaken off (*str'akhnut'*), blown away (*sdut'*), and scattered (*razveyat'*), but only *zola* can be raked apart (*razgresti*) and scooped out (*vygrebat'*), while both can be raked (*razgrebat'*); one can be reborn (*vozrodit'sya*) only from *pepel*, but bake (*ispech'*) something only in *zola*; only the lexeme *pepel* enters into coordinative constructions with *almaz* 'diamond', *okurok* 'cigarette butt', *lava* 'lava', and *razvalina* 'ruin', but only *zola* — with *navoz* 'manure', *golov'oshka* 'ember', and *sazha* 'soot'. Overall, the lexeme *zola* is used when the corresponding substance is used for something, cf. the following examples (cit. via RNC, translated by DeepSeek — S.B.):

"When worms attack this tree, one should water its stump with vinegar, or sprinkle it with ashes (*zola*)" (V. A. Levshin. Vseobshcheje i polnoje domovodstvo, part V, 1795);

"To destroy it — i.e., the earth flea — they sprinkle flax seedlings with ashes (*zola*), I read in one book" (A. N. Engelgardt. Pis'ma iz derevni. Pis'mo tretje, 1872);

"They removed the worm, sprinkled the cabbage with ashes (*zola*) against the fleas, watered the garden" (N. A. Leykin. Derevenskaja prelestnitsa, 1908);

"I fumigate the field with smoke, sprinkle it with ashes (*zola*) from a basket" (I. A. Aramilev. V lesakh Urala. *Novyj mir*, 1941);

"Burdenko pressed against someone's gates and watched excitedly as she walked along the high wooden sidewalk, sprinkled with stove ashes (*zola*) so it wouldn't be slippery" (Pavel Nilin. Interesnaja zhizn', 1969–1980);

"In the valleys of the Pamirs, local residents have long sprinkled fields with ashes (*zola*) and earth to accelerate spring snowmelt" (L. R. Serebr'annyj, A. V. Orlov. Ledniki v gorax, 1985);

"And now it was enough to eat, say, a sprat, drink, say, some Cahors, catch a cold, moving between snowdrifts along paths sprinkled with ashes (*zola*) in the settlement's streets, and the old ailment made itself known, weakly, quietly, but noticeably" (Asar Eppel. Poka i poskol'ku, 1991);

"All cuts, especially on roots, are sprinkled with wood ashes (*zola*) — roots dusted with ashes (*zola*) receive an additional stimulus for development" (Valerija Iljina. Pokor'ajushchije vysotu. *Nauka i zhizn'*, 2009);

"To obtain potash, dense wood was burned to ashes (*zola*), the ashes (*zola*) were boiled in water for several days in specially adapted copper pots" (Arkadij Kuramshin. Elementy. Zamechatel'nyy son professora Mendeleeva, 2019).

If we are talking about *pepel*, the contexts are completely different: besides the collocations listed above, *pepel* is primarily what is obtained as a result of burning something or someone, cf. the following examples (cit. via RNC, translated by DeepSeek — S.B.):

"For that, considerable booty was obtained, and the whole village was turned into ashes (*pepel*)" (F. I. Soymonov. Opisanije Kaspijskogo mor'a, 1763);

"Do not burn me to ashes (*pepel*) with your anger! " (A. A. Bestuzhev-Marlinskij. Mulla-Nur, 1836);

"He was cast out of a window, massacred, burnt, and his ashes (*pepel*) were blown abroad from a cannon's mouth, to the four winds of heaven!" (A. S. Pushkin. Kapitanskaja dochka, 1836);

"... she only turned pale, hastily burned the letter and blew its light ashes (*pepel*) onto the floor" (M. Yu. Lermontov. Kn'agin'a Ligovskaya, 1836–1837);

"His path was strewn with corpses and ashes (*pepel*) of burned villages and cities" (N. N. Alekseev. Brat na brata, 1904);

"When the hut burned down, the bravest ventured to rummage in the ashes (*pepel*)..." (Yevgenij Vodolazkin. Lavr, 2012);

"There among the burned corals and ashes (*pepel*), Albert Ghiorso discovered atoms of element number 99..." (Arkadij Kuramshin. Elementy. Zamechatel'nyy son professora Mendeleeva, 2019).

Even when *pepel* is used for something, it is necessarily indicated what exactly was burned (for *zola* such an indication is not obligatory), cf. (cit. via RNC, translated by DeepSeek. — S.B.):

"After that, this earth is covered with a smaller layer of mixed matter, which contains two parts of ashes (*pepel*), burned from hardwoods, and three parts of quicklime..." (M. V. Lomonosov. O varenii selitry, 1741);

"If he was strong enough, he uprooted stumps, burned the whole pile of felled wood and sowed bread on the earth fertilized with ashes (*pepel*), putting it under hay" (N. I. Berezin. Peshkom po karel'skim vodopadam, 1903).

Furthermore, *pepel* is used in metaphorical contexts to a greater extent than *zola*, cf., for example (cit. via RNC, translated by DeepSeek. — S. B.): "A dry face. Eyes extinguished, as if sprinkled with ashes (*pepel*)" (G. Ya. Baklanov. Pyad' zemli, 1959).

Also, *pepel* (but not *zola*) denotes such a specific thing as the gray coating that forms on the surface of just extinguished coals, cf. (cit. via RNC, translated by DeepSeek. — S. B.):

"...a coal that had just gone out, covered in places with *pepel*..." (F. P. Wrangel. Puteshestvije po severnym beregam Sibiri i po Ledovitomu mor'u, 1841);

"Despite the hot season, embers covered with *pepel* glowed in the fireplace" (F. F. T'utchev. Na skalax i dolinax Dagestana, 1903).

Thus, apparently, speakers of Russian choose the lexeme *pepel* when they want to emphasize that these are the remains of something (or someone) burned, and the lexeme *zola* when they want to emphasize that the corresponding substance is used for a specific purpose. Interestingly, it was the Russian word *zola* that was chosen as the main word for the slot 'ashes' in Kassian et al. 2010, despite the fact that the lexeme *pepel* has a higher frequency, and this decision appears to be well justified.

Incidentally, it is possible that the situation with PIE **péh$_2$wr̥* and **h$_1$n̥g$^{w}$nís* 'fire' was also of this sort. Since these two nouns display different genders (**péh$_2$wr̥* is neuter, while **h$_1$n̥g$^{w}$nís* is masculine), they may have originally referred to 'fire' as 'substance' and 'fire' as a 'personal' entity respectively (see Yakubovich 2021; Schmidt 2010 etc.). An anonymous reviewer suggests that 'fire as substance' would be a more natural candidate for lexicostatistical purposes.

Apparently, distinctions of this sort are further evidence of the absence of any presumed innate universal, language-independent mental lexicon suggested in some works — such as "mentalese" (Pinker 1994) or "universal object code" (Zhinkin 1998). Each language at each moment in time clusters the infinite variety of elements of surrounding reality in different ways and on different grounds. Certainly, some similar clustering options can be found quite often, but at any moment the next language available for research can draw boundaries between concepts in its own way.

The differences in ways of clustering *realia* are now actively studied within lexical typology (see, e.g., Reznikova et al. 2020). Perhaps someday a list of possible semantic distinctions, completely exhausting the possibilities of the human mind, will be compiled. But until this is done, it probably makes sense to adapt the methodology so that it can work with lists that take into account the possibility of "synonyms" (such that the real differences between their meanings are not removed by the suggested contexts) appearing at any stage of reconstruction.

That said, these observations should not be interpreted as a call to allow any number of synonyms for any word in the wordlist of any proto-language at any stage of reconstruction. For most of the described processes, a large amount of time must pass: for instance, unrecog-

nizable synonymy which arises due to phonetic or morphological reasons usually requires a chain of phonetic changes, including potentially irregular ones, that obscure the original phonemic composition, and the number of daughter branches must be minimal. (Perhaps this is why language isolates, which have no close relatives, in many cases cannot be convincingly assigned to any macro-family.) When a language does have close relatives, the methodology of step-by-step reconstruction (involving the maximum available amount of data), such as the one adopted by the Moscow School of comparative linguistics, in most cases makes it possible to determine which of the synonyms was the primary one in the proto-language or, at least, could be considered an optimal choice next to any competition (Starostin 2010: 102).

Nevertheless, all of this means that it is always necessary to accompany any reconstructions in onomasiological databases with comments listing their potential quasi-synonyms with references to semantic distinctions that were already tested and excluded (thus, when a new semantic distinction is found, it will be easy to see which reconstructed lexemes need to be retested). In those cases where it is impossible to identify one primary candidate, it makes sense to include all of them into reconstructed wordlists, assigning equal (or unequal) probabilities to each one. This would make it easier to reexamine them when a new semantic distinction (or additional sociolinguistic information) comes to light.

## References


Belikov, Vladimir. 2006. *Konvergentnyje protsessy v lingvogeneze (Convergence in linguogenesis)*. Doctoral dissertation: Moscow.

Black, Paul. 1997. Lexicostatistics and Australian languages: Problems and prospects. In: Darrell T. Tryon, Michael Walsh (eds.). *Boundary rider: essays in honour of Geoffrey O'Grady*. *Pacific linguistics, C-136*: 51–69. Canberra: The Australian National University.

Black, Sheila R. 2003. Review of Semantic Satiation. In: S. P. Shohov (ed.). *Advances in psychology research*. *Vol. 26*: 63–74. Nova Science Publishers.

Burlak, Svetlana. 2021. Ustoichivost' i chastotnost': jest' li korrel'atsija? (Stability and frequency: is there a correlation?). *Journal of Language Relationship* 19(3–4): 293–307.

Campbell, Lyle. 2004. *Historical Linguistics: An Introduction. 2nd edition*. Cambridge: The MIT Press.

Dunn, Michael John. 1999. *A Grammar of Chukchi*. Ph.D. thesis: Australian National University.

Dyakonov, Igor'. 1984. Sravnitel'noe jazykoznanije, istorija i drugije smezhnyje nauki (Comparative linguistics, history and neighboring sciences). In: *Lingvisticheskaja rekonstruktsija i drevnejshaja istorija Vostoka. Tezisy i doklady konferentsii. Part 2*: 3–20. Moskva: Institut vostokovedeniya RAN.

Dyachenko, S. V., A. V. Malysheva, A. V. Ter-Avanesova. 2018. Iz nabl'udenij nad fonetikoj vladimirsko-povolzhskix govorov: jokan'je i pozitsionnaja m'agkost' *zh*, *sh* (Some observations on the Vladimir-Povolzh'ye dialect phonetics: yokan'ye and positionally determined palatal articulation of *zh* and *sh*). *Trudy Instituta russkogo jazyka imeni V. V. Vinogradova* 3(17): 41–68.

Frankham, Richard, C. J. A. Bradshaw, B. W. Brook. 2014. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation* 170: 56–63.

GLD 2025 = Starostin, George (ed.). 2011–2025. *The Global Lexicostatistical Database*. Moscow: Higher School of Economics, & Santa Fe: Santa Fe Institute. Available online at: https://starlingdb.org/new100/ [accessed September 1, 2025].

Harcourt, Alexander H. 2002. Empirical estimates of minimum viable population sizes for primates: Tens to tens of thousands? *Animal Conservation Forum* 5(3): 237–244. doi:10.1017/S1367943002002287

Haspelmath, Martin, Uri Tadmor (eds.). 2009. *Loanwords in the world's languages: a comparative handbook*. Berlin: Walter de Gruyter GmbH & Co.

Henrich, Joseph. 2016. *The secret of our success. How culture is driving human evolution, Domesticating our species, and making us smarter*. Princeton University Press.

Holmes, Janet, Nick Wilson. 2022 *An introduction to sociolinguistics*. *6th edition*. London: Routledge.

Iomdin, Boris. 2023. *Telegram-kanal «Uznal novoe slovo»* (*Telegram channel "I've got to know a new word"*). Available online at: https://t.me/UznalSlovo/627.

Jackson, Matthew O. 2019. *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors*. New York: Knopf Doubleday Publishing Group.

Kassian, Alexey, George Starostin, Anna Dybo, Vasily Chernov. 2010. The Swadesh wordlist. An attempt at semantic specification. *Journal of Language Relationship* 4: 46–89.

Khrisanfova, Elena, Ilya Perevozchikov. 2005. *Antropologija* (*Anthropology*). Moskva: Nauka.

Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog, Seongha Rhee. 2019. *World lexicon of grammaticalization*. *2nd ed*. Cambridge University Press.

Malyutina, S. A., O. V. Dragoy, A. G. Petrushevskiy, O. N. Fedina, M. V. Ivanova, D. A. Sevan, E. F. Gutyrchik. 2012. Nejrofiziologicheskije korrel'aty chastotnosti slova pri jego nazyvanii (Neurophysiological correlates of the word frequency while citing it). In: *Mezhdunarodnyj simpozium po nejroimidzhingu: fundamental'nyje issledovanija i klinicheskaja praktika*: 83–85. Moskva: MGPPU.

Milroy, Lesley, Matthew Gordon. 2008. The concept of social network. In: Lesley Milroy, Matthew Gordon. *Sociolinguistics: Method and interpretation*: 116–133. Oxford: John Wiley & Sons.

Moulin, Chris J. A., Nicole Bell, Merita Turunen, Arina Baharin, Akira R. O'Connor. 2021. The induction of jamais vu in the laboratory: word alienation and semantic satiation. *Memory* 2021 Aug. 29(7): 933–942. doi: 10.1080/09658211.2020.1727519.

Normanskaya, Julia V. 2005. *Razvitije i genezis sistem tsvetooboznachenij v drevnix indojevropejskix jazykax* (*Development and genesis of the color name systems in ancient Indo-European languages*). Moskva: ZAO «S&K».

Ossovetskiy, I. A. (ed.). 1969. Slovar' sovremennogo russkogo narodnogo govora (d. Deulino R'azanskogo rajona R'azanskoj oblasti (Dictionary of a modern Russian dialect: Deulino village, Ryazan' region). Moskva: Nauka.

Pinker, Steven. 1994. *The Language Instinct: How the Mind Creates Language*. New York: William Morrow and Company.

Pogibenko, Tamara. 2013. *Avstroaziatskije jazyki: problemy grammaticheskoy rekonstruktsii* (*Austro-Asiatic languages: the problems of grammar reconstruction*). Moskva: Institut vostokovedenija RAN.

Pogibenko, Tamara. 2018. Leksicheskij kompleks v jazyke ma (Lexical complex in the Ma language). *Vestnik Instituta vostokovedenija RAN* 5: 200–211.

Pozdnyakov, Konstantin. 1993. *Sravnitel'naya grammatika atlanticheskix jazykov* (Comparative grammar of the Atlantic languages). Moskva: Nauka.

Polivanova, Anna. 2013. *Staroslav'anskij jazyk. Grammatika. Slovari* (*Old Church Slavonic language. Grammar. Dictionaries*). Moskva: Russkij fond sodejstvija obrazovaniju i nauke.

Reznikova, T. I., E. V. Rakhilina, D. A. Ryzhova. 2020. Glagoly padenija v jazykax mira: frejmy, parametry i tipy sistem (Verbs of falling in the world's languages: frames, parameters and system types). *Acta Linguistica Petropolitana* 16(1): 9–63. doi: 10.30842/alp2306573716101

RNC = 2003–2025. *The Russian National Corpus*. Available online at: ruscorpora.ru [accessed on September 1, 2025].

Samarina, Irina. 2007. Protsessy monosillabizatsii v jazykax materikovoj Jugo-Vostochnoj Azii (Monosyllabization processes in the languages of continental Southeast Asia). *Voprosy filologii* 1(25): 41–49.

Schmidt, Karl Horst. 2010. Miscellanea Anatolica. In: J. Klinger et al. (eds.). *Investigationes Anatolicae: Gedenkschrift für Erich Neu*: 241–252. Wiesbaden: Harrassowitz.

Simons, Gary F. 1982. Word taboo and comparative Austronesian linguistics. In: Amran Halim, Lois Carrington, S. A. Wurm (eds.). *Papers from the third international conference on Austronesian linguistics, Vol. 1: Currents in Oceanic*. *Pacific linguistics, C-74*: 157–226. Canberra: The Australian National University.

Skorik, P. J. 1961. *Grammatika chukotskogo jazyka (Grammar of the Chukchi language)*. Moskva / Leningrad: Izdatel'stvo Akademii nauk SSSR.

Smirnitskaya, Anna. 2024. *Dravidijskie terminy rodstva: tipologija semanticheskix perexodov* (*Dravidian kinship terms: the typology of semantic shifts*). Moskva: Institut vostokovedenija RAN.

Starostin, George. 2010. Preliminary lexicostatistics as a basis for language classification: A new approach. *Journal of Language Relationship* 3: 79–116.

Starostin, George. 2019. Chinese basic lexicon from a diachronic perspective: implications for lexicostatistics and glottochronology *Journal of Language Relationship* 17(2): 153–176.

Sumbatova, Nina. 2022. Landuma: a case of radical alliterative agreement. *Journal of African Languages and Linguistics* 43(1): 83–118. doi:10.1515/jall-2022-8893.

Tadmor, Uri. 2015. When culture grammaticalizes: The pronominal system of Onya Darat. In: R. De Busser, R. J. LaPolla (eds.). *Language Structure and Environment. Social, cultural, and natural factors*: 77–98. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Timoshenko, S. P., L. L. Iomdin, S. A. Gladilin, E. S. Inshakova. 2021. SinTagRus v sostave NKRJa: novyje vozmozhnosti (SynTagRus within the Russian National Corpus: New Possibilities). In: V. P. Zaxarov (ed.). *Korpusnaya lingvistika — 2021. Trudy mezhdunarodnoj konferentsii*: 31–43. Sankt-Peterburg: Skifija-Print.

Vallejos-Yopán, Rosa. 2015. La indexicalidad de género en kukama-kukamiria desde una perspectiva tipológica. In: Zarina Estrada Fernández, Ana Fernández Garay, Albert Alvarez Gonzalez (eds.). *Estudios de lenguas amerindias 3: Escenarios de diversidad lingüística*: 195–222. Hermosillo, Sonora: UNISON.

Vavilov, Nikolay. 1935. *Zakon gomologicheskix r'adov v nasledstvennoj izmenchivosti* (*The law of homology series in genetical mutability*). Moskva/Leningrad: Sel'xozgiz.

Vlasova, R. M., V. E. Sinitsyn, E. V. Pechenkova. 2015. The Effect of Word Frequency on the Brain Correlates of Object Naming in Russian. *The Russian Journal of Cognitive Science* 2(1): 24–40.

White Hat, Albert. 1999. *Reading and writing the Lakota language*. Salt Lake City: University of Utah Press.

White, Ralph Gardner. 1968. Borrowing and Taboo in Eastern Polynesia. *The Journal of the Polynesian Society* 77(1): 57–73.

Yakubovich, Ilya. 2021. Hittite agent form. In: A. I. Kogan, A. S. Panina (eds.). *Problemy obshchej i vostokovednoj lingvistiki. Sochetaemost' jazykovyx jedinits i jazykovyje modeli. Pam'ati Z. M. Shal'apinoj (1946–2020)*: 124–141. Moskva: Institut vostokovedenija RAN.

Yelisejeva, M. B., E. A. Vershinina, V. L. Ryskina. 2017. *Makarturovskij oprosnik: russkaja versija. Otsenka rechevogo i kommunikativnogo razvitija detej rannego vozrasta. Normy razvitija. Obraztsy analiza. Kommentarii* (*The MacArthur Communicative Development Inventories: Russian version. Assessment of speech development in infants. Development norms. Examples of analysis. Comments*). *2nd edition, revised and augmented*. Ivanovo: LISTOS.

Zhinkin, Nikolay. 1998. *Jazyk — rech — tvorchestvo: Issledovanija po semiotike, psixolingvistike, poetike* (*Language, speech, creation: Research in semiotics, psycholinguistics, poetics*). Moskva: Labirint.

*С. А. Бурлак*. О возможных источниках лексических замен и синонимии в праязыковых списках базисной лексики

При ступенчатой реконструкции встречаются ситуации, когда на одно и то же место в протосписке претендует более одного слова, и при этом нет возможности отвести один из потенциальных синонимов стандартными способами. Вместе с тем, в реально наблюдаемых языках практически всегда можно обосновать выбор слова для включения в список. Трудности проведения такого обоснования для праязыков могут быть связаны как с социолингвистическими, так и с внутриязыковыми причинами. Поскольку нам неизвестна социальная картина праязыка, а также её эволюция в дочерних популяциях, может не хватить материала для выводов о том, что слово, которое в праязыке было ограничено какими-то социальными ситуациями (например, употреблялось только одним из гендеров, или только по отношению к выше/ниже стоящему в иерархии, или только в разговоре с детьми и т. п.), могло при изменении общественного уклада стать основным в какой-то из дочерних ветвей. К увеличению числа синонимов в языке ведёт практика табу: в каждый конкретный момент основным является одно из слов, но в разные моменты это может различаться, так что одни из дочерних популяций унаследуют в качестве основного одно слово, другие — другое. Кроме того, в дочерних ветвях в разные стороны могли пойти фонетические или морфологические выравнивания, так что основы, прежде составлявшие одну (возможно, супплетивную) парадигму, стали единственными представителями соответствующего значения в каждой из дочерних ветвей. Чем больше времени прошло от разделения праязыка, тем труднее обнаружить те чередования, которые вызвали первоначальное различие между наблюдаемыми формами, выглядящими как несводимые друг к другу. Возможно, главная труд-

ность для выделения основного слова состоит в том, что в праязыке могла быть иначе устроенная номинационная решётка, так что те контексты, которые нынешнему наблюдателю представляются однородными, стали считаться таковыми лишь во время существования языков-потомков, при этом одни из них в качестве основного взяли одно из слов, другие — другое.

*Ключевые слова*: стословный список Сводеша; синонимия; лексические замены; сравнительно-исторический метод; семантическая реконструкция; языковое родство; лексикостатистика.