Significance testing of the Altaic family

Andrea Ceolin

University of Pennsylvania

Copyright: *Diachronica* 36:3 (2019), pp. 299–336
issn 0176-4225 | e-issn 1569-9714
https://doi.org/10.1075/dia.17007.ceo

Abstract
Historical linguists have been debating for decades whether the classical comparative method provides sufficient evidence to consider Altaic languages as part of a single genetic unity, like Indo-European and Uralic, or whether the implicit statistical robustness behind regular sound correspondences is lacking in the case of Altaic. In this paper I run a significance test on Swadesh-lists representing Turkish, Mongolian and Manchu, to see if there are regular patterns of phonetic similarities or correspondences among word-initial phonemes in the basic vocabulary that cannot be expected to have arisen by chance. The methodology draws on Oswalt (1970), Ringe (1992, 1998), Baxter & Manaster Ramer (2000) and Kessler (2001, 2007). The results only partially point towards an Altaic family: Mongolian and Manchu show significant sound correspondences, while Turkish and Mongolian show some marginally significant phonological similarity, that might however be the consequence of areal contact. Crucially, Turkish and Manchu do not test positively under any condition.[1]

Keywords: comparative method, historical linguistics, Altaic, lexicostatistics, Swadesh lists, multilateral comparison

1 Introduction

Traditional Altaicists (Ramstedt 1957; Poppe 1960, 1965; Menges 1975; Manaster Ramer & Sidwell 1997) and Nostraticists (Bomhard 1996, 2008, 2011; Dolgoposky 1986; Illič-Svityč 1971; Starostin 1991), and in particular Starostin et al. (2003), have argued that sound correspondences among Turkic, Mongolic and Tungusic can be identified through a rigorous application of the classical comparative method. For some scholars, the correspondence schemes can be extended to Korean and Japanese, see Miller (1971, 1996) and especially Robbeets (2005). On the other hand, many Altaic specialists have criticized the methodology behind the identification of such correspondences, and are skeptical about the genetic relatedness of Altaic languages (Doerfer 1973; Sinor 1988; Unger 1990; Georg 1999; Vovin 2005).

The main question driving the debate is how the classical comparative method should be applied in cases in which we might expect some difficulties in retrieving regular sound

[1] Supplementary material available at http://github.com/AndreaCeolin/Significant-Testing-of-Altaic.

correspondences: for instance, if we try to compare languages that have been separated for several millennia, and therefore cognate words might be irretrievable, or if the lack of historical records makes it challenging to reconstruct a clear set of correspondences, or to distinguish between possible cognates and loanwords.

A promising research line has focused on developing statistical methods to distinguish patterns of similarities or correspondences that can be judged significant, under hypothesis testing, from patterns that can be expected to arise simply by chance, once one takes into account the number of comparanda and the phonemic distribution of the languages under investigation (Ringe 1992, 1998; Baxter & Manaster Ramer 2000; Kessler 2001; Kessler & Lehtonen 2006).

In this paper, I apply some statistical methods proposed in the literature to look for significant patterns of similarities and correspondences in a restricted list of words belonging to Turkish, Mongolian and Manchu, as representatives of the three families that constitute the most narrow hypothesis of Altaic. In §2, some background information about the current state of the debate over Altaic is provided. Section 3 describes some of the methods proposed for long-range comparison. Section 4 explains how wordlists have been collected. Section 5 introduces the problem of polymorphism. Section 6 contains a first test on languages whose genetic relation is established and languages which are not provably related. In §7, the tests are applied to the three Altaic languages under examination. Finally, §8 addresses the robustness of the results with respect to the possible presence of loanwords.

2 The Altaic controversy

It has been clear for decades that the evidence in support of the Altaic family is controversial, especially if compared to that in support of other established families, like Indo-European and Uralic. Comrie (1981: 40) summarizes the main problems in the evaluation of the most restrictive Altaic hypothesis (i.e., Turkic, Mongolic and Tungusic): i) few features are common to the three postulated branches: in fact, while shared vocabulary and morphology can be found between Mongolic and Turkic, and between Mongolic and Tungusic, the evidence between Tungusic and Turkic is more scarce; ii) one of the strongest parallels among the three groups, i.e. personal pronouns, is also present among other non-Altaic languages of Eurasia (Indo-European, Uralic, and Chukotko-Kamchatkan), and therefore can be used either as an argument for areal contact or as an argument for a larger Eurasiatic macrofamily, but not for Altaic; iii) the agglutinative nature of Altaic morphology makes it less reliable as a source of genetic relatedness, because it can be subject to borrowing.

One of the main sources of evidence in favor of the Altaic hypothesis comes from the *Etymological Dictionary of the Altaic Languages* (EDAL, Starostin et al. 2003), which contains an exhaustive list of etymologies that have been proposed in support of the Altaic family. However, some scholars have criticized the methodology employed by the authors. Vovin (2005) lists in detail many potential flaws in Starostin et al.'s specific approach to the historical reconstruction of Proto-Altaic (PA): i) the evidence for the family comes mostly from vocabulary lists rather than from morphology, traditionally considered a more reliable source of historical information. The only morphological evidence Starostin et al. rely on is derivational morphology, more subject to borrowing, while inflectional morphology is mostly ignored; ii) the reconstruction of PA words is often in contradiction with the internal reconstructions of the proto-languages of the single families; iii) some PA words are only attested in a single language, and sometimes in a single dialect; iv) the semantic classes to which cognate words are ascribed are vague or, even worse, stretched to fit the reconstruction; v) the proposed phonetic correspondences have many exceptions and the same phoneme can be associated with many different rules, to the extent that it is difficult to distinguish such correspondences from accidental similarities.

On the other hand, Dybo & Starostin (2008) cite many cases of well-established families, like Indo-European and Austronesian for which some of Vovin's criticisms would apply as well. A clear example is the vowel correspondences between German and English (Dybo & Starostin 2008: 146-147) which, the authors argue, would be, without historical or dialectal information, by no means more convincing than the vowel correspondences among the Altaic languages, given the high level of phonological conditioning and exceptions.[2] Overall, it is interesting to note how Dybo & Starostin's reply is more focused on showing how criticisms of the Altaic reconstruction can be applied to well-established families as well, rather than directly addressing problems identified by Vovin.

Perhaps the most convincing reaction to Vovin's criticisms is represented by Robbeets (2005). Even though she focuses mostly on identifying cognates between Japanese and the other Altaic languages, including Korean, without entering the debate on the core of the Altaic family, her scrutiny of the etymologies proposed in the literature provides evidence for some clear sound correspondence schemes. In the selection of cognate words, Robbeets excludes etymologies that contradict internal reconstructions, are poorly attested or whose semantic match is not clear, therefore excluding most of the classical etymologies and focusing on 359 'core' items. Furthermore, Robbeets (2015) also attempts a comparison of verb morphology that shows how one can also find evidence for the Altaic family looking outside of the basic vocabulary. Even though Robbeets's work aims at reconstructing a larger Altaic family (including Korean and Japanese), it is clear that her methodology and her reconstructions represent the best evidence so far in favor of the Altaic hypothesis.

This paper mostly focuses on Vovin's point (v), which is a problem that Robbeets only partially addresses. While she recognizes that multilateral comparison decreases the significance of correspondence sets exponentially (Robbeets 2005: 286-287), her calculations do not take into account factors like vocabulary length and phonemic distribution, which are crucial to evaluate the likelihood of a correspondence set (Ringe 1992). Moreover, Georg (2008) argues that the semantic criteria used by Robbeets to select cognate words are not consistent, even though it is widely acknowledged that establishing strict criteria of semantic matching is not an easy enterprise, given that semantic shift is a natural form of language change. Both problems will be addressed in the methodology presented here.

3 Methods

In this paper, I use wordlist comparison to see if any trace of phonetic similarities or correspondences in support of Altaic can be retrieved by looking at the basic vocabulary of some modern Altaic languages. There are several reasons why I decided to focus on the basic vocabulary rather than on other domains, for instance paradigmatic morphology.

[2] A similar argument can be found in Comrie (1998) involving the comparison of Old English and Modern English pronouns: the number of exceptions in this restricted domain is considerable.

First, Nichols (1996) argues that determining probabilistic thresholds for significance testing relying on paradigmatic morphology is more difficult with agglutinative languages like Altaic than with inflected languages like Indo-European, because the chance of horizontal transmission cannot be excluded.

Second, even though morphological evidence is less controversial than evidence coming from basic vocabulary, because suffixes are more resistant to borrowing, morphology is not immune from contact. For instance, loss of case-marking and case reanalysis can be contact-induced changes. Moreover, Dybo & Starostin (2008: 125-126) argue that if we built an exhaustive list of Indo-European morphological suffixes, the percentages of those which could be tracked directly from modern languages to Proto-Indo-European would be small compared to the number of Indo-European roots we can retrieve from traditional Swadesh lists. This means that while morphological evidence is more probative than lexical evidence, a researcher is very likely to encounter situations where the morphological signal has been completely lost while the lexicon still conserves some properties of the proto-language.[3]

Third, morphological suffixes are usually shorter than words, and therefore one needs to rely on a large number of items to reach individual-identifying evidence for relatedness, in terms of Nichols (1996). Finally, the literature on wordlist comparison is vast and provides an excellent background for discussing methods, problems and results related to long-range comparison (Ross 1950; Swadesh 1955; Greenberg 1957; Oswalt 1970; Villemin 1983; and the articles in Salmons & Joseph 1998 and Kessler 2001 among many others). In particular, Kessler's works are the most exhaustive when it comes to methodological discussions and statistical design, and therefore most of this paper will draw from these.

Furthermore, I decided to make the unconventional choice of focusing on modern languages rather than reconstructed proto-forms. The main reason for this methodological choice is to avoid the problems of significance tests as the one presented in Kassian et al. (2015), who argued in favor of an Indo-Uralic macrofamily on the basis of phonetic similarities in the basic vocabulary of Proto-Indo-European and Proto-Uralic. The paper shows there are some degrees of freedom in the selection of reconstructed proto-forms, even in a well-studied domain like Indo-European. Determining whether a consonant should be reconstructed or not for the proto-language, or whether it is part of the root or it is a morpheme, are all decisions with a high impact on the data selection and on the interpretation of the results, as noted in Kessler (2015) and Ringe (2015). Scholars usually disagree on the legitimacy of the forms to include, which is dependent on the hypothesis of reconstruction. The choice of the number of proto-forms to reconstruct and the inevitable uncertainty associated with their semantics introduce other human biases in the wordlist. For these reasons, I think that focusing on modern varieties is a safer choice: it reduces the risk of human biases, now limited to language selection and the match between words and semantic classes, and it allows researchers to discuss methods and results that are not heavily dependent on the availability of historical documents or of a long tradition of philological studies. These are both factors that facilitate the process of historical reconstruction, but are mostly limited to families which have already been independently established by the classical comparative method.

Of course, this choice has the consequence of reducing the occurrence of Type I errors (false positives), i.e. of finding evidence for a genealogical claim because the data is biased

[3]For an attempt to run a test relying on evidence coming from morphosyntactic characters, see Longobardi et al. (2013, 2016)

against the null hypothesis, but it runs the risk of producing a test which is too conservative, i.e. a test that yields many Type II errors (false negatives). We will see what implications the methodology has for the comparison of some Indo-European languages as a case study in §6. The next subsections describe the methods employed in this paper.

3.1 Phonetic algorithms

One class of methods focuses on phonetic similarities between wordlists rather than sound correspondences, with the motivation that the latter might be hard to identify after a certain time span. These methods are generally inspired by Oswalt (1970), and the one employed in this study is the Monte Carlo method described in Kessler & Lehtonen (2006).

Their algorithm runs on vocabulary lists and focuses on the place of articulation of the first consonant. Consonants are divided into five classes, and a distance measure between two words is calculated according to the index assigned to the class of their word-initial consonants: labial (0), anterior (4), palatal (6), velar (9) and postvelar (10).

In order to determine whether two languages are phonetically similar in a significant way, a global phonetic distance is calculated between two aligned word-lists. Afterwards, the same procedure is repeated permuting the order of the words in either list, i.e. breaking the word-meaning association, and a new phonetic distance is then calculated.

If we repeat the second procedure enough times, we obtain a simulated distribution which provides us with a reasonable estimate of which kind of distance we would expect to find by chance given the distribution of the phonemes in the two languages. This allows us to determine whether the phonetic distance between two languages is significantly small or not.

Other properties of Kessler & Lehtonen's algorithm are: i) in the case of polymorphism, namely the fact that languages can use multiple words to express the same meaning, the phonetic distance is the average of the values that are obtained taking into account all the possible combinations of the words in the two languages. For instance, if a meaning has two words in each language, all the four possible distances are calculated and the average is returned; ii) in case of double articulation, all the combinations of the sounds are considered as well, but the smaller distance is chosen instead. iii) non-lexical words, like pronouns and functional words, are removed, because they are expected to be redundant and less arbitrary.

The algorithm can be easily modified to handle multiple comparison, but since in this paper we focus only on three languages, the tests will be pairwise.

3.2 Phonological classes

A second class of methods focuses instead on correspondences between phonological classes rather than phonetic similarities. The best representative of this class is the algorithm proposed in Baxter & Manaster Ramer (2000) and then again in Kessler (2007), which adopts the same Monte Carlo procedure described in Kessler & Lehtonen (2006). The main difference is that the method relies on Dolgopolsky's phonological classes (Dolgopolsky 1986), which are grouped in the following categories:

- Labial Obstruents (/p/, /f/, /v/)
- Dental or Apical Obstruents (/t/, /d/)
- Sibilants (/s/, /ʃ/)

- Palatal, Dorsal, Postalveolar affricates (/k/, /g/, /tʃ/ /dʒ/)
- Labial Nasal /m/
- other Nasals (/n/)
- Liquids (/r/, /l/)
- Rounded semivowels (/w/ and word-initial /u/)
- Palatal Approximant (/j/)
- Vowels + Dorsal Nasals + Glottals (/h/, /o/, /ŋ/)

Since there are groups of phonemes that are likely to change into one other, the idea behind Dolgopolsky's classes is to group them into a single class, relaxing in this way the criterion for judging two sounds similar. With this metric the distance between two phonemes is either 1 or 0, depending on whether they belong to the same class or not, and the main difference with the previous metric is that it is applied to the first phoneme of the word rather than the first consonant. This means that word-initial vowels have their class as well.

Polymorphic characters are handled with the same strategy as Kessler and Lehtonen's algorithm.

3.3 Regular Correspondences

A third class of methods focuses instead on automatically detecting sound correspondences and identifying a threshold for which a number of correspondences should be considered significant. The relevant works are Ross (1950), Villemin (1983), Ringe (1992, 1998) and Kessler (2001). Ringe (1998) proposes the hypergeometric distribution as a way to estimate, given two lists of words, the likelihood of finding a number of correspondences equal to or higher than a certain value r for a given pair of sound. If we consider a wordlist of length N, with a sound occurring n times in language A, and another sound occurring R times in language B, we can calculate the hypergeometic random variable h:

$$h = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}$$

The distribution obtained changing the parameter r is hypergeometric. Calculating the cumulative distribution associated with the number of r correspondences identified, by summing the various h values, allows us to estimate how likely it is that a sound correspondence results from chance. According to Ringe, a sound correspondence is significant if the likelihood of its occurrence, given by the cumulative distribution, is lower than .01. This method can be used to determine if the number of sound correspondences automatically identified between two wordlists is sufficiently different from the number of correspondences that one might expect to find by chance, given the length of the wordlist and the initial phoneme distribution in the languages, by means of the same Monte Carlo tests applied in the previous methods. As the previous methods, its simplest application uses just the first phoneme of the word as a data point.

The main problem of this method is that, according to Ringe (1998), it is not immune to false positives and false negatives. Baxter (1998) and Kessler (2001) show that because of the nature of the hypergeometric distribution, single matches in rare consonants can be as significant as correspondence sets drawn from more common phonemes. This produces, inevitably, some false positives when real languages happen to exhibit a match between rare phonemes, or false negatives when an established set of correspondences between two related languages is not judged sufficient after a permutation test because other sets of correspondences, supported by a smaller number of phonemes, appear after the lists are randomized.

Kessler (2001) addresses the problem by developing an alternative metric, defined as R^2, which can be summarized in the following equation:

$$R^2 = \sum_1^n (n-1)^2$$

Where n is the number of correspondences for each phonemic match that exhibits at least one correspondence. The reasoning behind the metric is the following. It is intuitive that by just summing the number of total correspondences between the word initial phonemes of two wordlists, one would end up with a number equal to the length of the wordlists. For instance, two wordlists of 100 words each one would have exactly 100 different phonemic correspondences. Some of these correspondences, however, will involve the same phonemic match. Therefore, this sum would be different if we subtract 1 from every single attested match: in this way, pairs of languages where there are many singletons (i.e., phonemic matches with exact one correspondence) will result in a lower sum with respect to pairs where phonemic matches tend to have either zero or more than one instance, a scenario that we would predict in case of language relatedness. This sum can be therefore used to estimate the presence of true correspondences between two languages. Furthermore, by squaring the count, the weight given to those phonemic matches supported by several correspondences increases exponentially.

Kessler (2001) shows that this metric is less sensitive to false positives and false negatives than Ringe's hypergeometric test. For these reasons, the R^2 metric will be used to evaluate sound correspondences across wordlists by means of the same permutation tests described in the preceding sections, along with Ringe's test.

As for the problem of polymorphism, the simplest strategy is to add multiple correspondence sets for one meaning if the meaning is polymorphic. For instance, when a meaning is polymorphic in both languages (e.g., both languages have two words per meaning), all the possible (four) correspondences are added.

4 Wordlists

I use three wordlists representing Turkish, Mongolian and Manchu as input for the algorithms. These languages represent the three branches that constitute the most restrictive Altaic hypothesis, and therefore the one that should be easier to test.

A previous experiment similar to the one presented here was performed in Oswalt (1998). Oswalt used several wordlists to evaluate phonetic similarities among a selection of languages ascribed to Nostratic. While the three languages under investigation in this paper were included in Oswalt's work, the wordlists were not available as an appendix to the paper, and so they could not be tested. Unfortunately, the lists collected in Starostin's database, Starling[4], cannot be used because, as made clear by Kessler & Lehtonen (2006) and Kessler (2007), the data collection must be unbiased from any kind of *a priori* hypothesis, while the wordlists in Starling explicitly assume a Nostratic framework. However, the database will be used as reference to evaluate specific cognates in §7.

For this reason, I decided to collect wordlists for the three languages from traditional dictionaries. For Turkish, I used Redhouse (1968), along with the judgments of a native speaker. For Mongolian, I used Hangin et al. (1986). For Manchu, I used the standard Norman (1978), and I cross-checked the entries with Rozycki (1994) and Li (2000).[5]

The starting point was a 207-Swadesh list, which results from the combination of the traditional 200- and 100-Swadesh lists (Swadesh 1955, 1971). Following the methodology in Kessler & Lehtonen (2006), 26 function words were removed, resulting in a 181-wordlist.[6] While I initially thought that this list would have represented a plausible baseline, an anonymous reviewer pointed out that there were clear cases in which the entries were influenced by onomatopoeia, especially in nursery words, or the words were clearly derived from other existing roots. For example, entries for the words *woman* (Mon. эм*(em)* , Man. *hehe*), *mother* (Mon. эх *(eh)*, Man. *eme*), *father* (Tur. *baba*, Man. *ama*) looked suspicious, and therefore the words were removed. The entries for *wife* and *husband* were also removed because of the risk of being derived from more basic words (cf. Mon. эхнэр *(ehner)* 'wife', a word that contains the root for 'mother'). An exception I made in this case was the entry for *male*, because while the Manchu word was a potential onomatopoeia word (*haha*, cf. also Robbeets (2005) about a similar entry in Japanese), Turkish and Mongolian display *er*, which was judged as a cognate in Starling.

A more delicate choice was whether to include the words for *breast* and *suck* (Tur. *göğüs* and *emmek*, Mon. хөх *(hoh)* and хөхөх *(hohoh)*, Man. *huhun* and *simimbi*), because Starling reports the set as a cognate set, but this is one of the typical words for which the risks of onomatopoeia and derivations from other words are very high; therefore, I decided to exclude them.

In addition, some other words looked like potential sources of redundancy:

- While the meaning 'tree' exhibits a traditionally accepted match between Mon. мод *(mod)* and Man. *moo*, the same root is used in Manchu in the word for 'stick'. Here one needs to make the methodological choice of either removing the Manchu entry or the entire meaning, but given that the word for 'stick' is likely to have a semantic overlap with other more common items, I decided to remove the entire meaning.

- Both Mon. хөл *(höl)* and Man. *bethe* are ambiguous between the meaning 'foot' and 'leg'. Therefore, the second meaning has been removed.

- A similar case was the word 'liver', which in Turkish has two reflexes derived from other words, *kara-* 'black' and *sakla-* 'to hide'. This was another case where rather than

[4]http://starling.rinet.ru

[5]I thank an anonymous reviewer for carefully checking the vocabulary entries and pointing me to some additional words that were not originally included in the lists, even though they satisfied the etymological criteria that I describe in the next section.

[6]The words are: *all, and, at, because, few, he, here, how, I, if, in, many, not, some, that, there, they, this, we, what, when, where, who, with, you*.SG, *you*.PL

removing the Turkish words, I decided to remove the meaning, since naming the organ with an already existing word is common practice (cf. the case of Romance, where for instance Italian *fegato* has the same root of English *fig* as a result of culinary practices).

- The meaning 'to kill' was also removed because of its semantic similarity with 'to die', which in Turkish shares the same root (Tur. *ölmek* 'to die' and *öldürmek* 'to kill').
- The meaning 'dust' was also removed because in both Turkish and Mongolian it shares its CV root with the word 'earth' (Tur. *toprak*, Mon. *товрог (tovrog)* 'dust' and Tur. *toz*, Mon. *тоос (toos)*.

These changes result in a final wordlist of 168 meanings.

5 Polymorphism

A common problem in dealing with wordlists is polymorphism: languages can use different words to express the same meaning. In the classical comparative method tradition, the problem of polymorphism has been largely ignored: if semantic shift changes the meaning of a word, in principle we can still prove that the word has a cognate, even though its meaning is different. For example, Eng. *clean* and Ger. *klein* are indisputable cognates, even though the meaning of the German word is different from the English one, because of a documented semantic shift. The cognacy judgment is possible because the evidence for the regular correspondence of the sounds in the two words is overwhelming in the vocabulary.

However, polymorphism becomes an issue when we want to evaluate a possible correspondence scheme through a statistical approach. If we choose more than one item to match a specific meaning, this must be taken into account in the statistical analysis. For these reasons, I decided to consider polymorphism only when the meanings are entirely overlapping according to the dictionary entries. I follow these criteria:

- Polymorphism in a meaning is not considered in cases in which the English word associated with the meaning is not listed among the possible translations. For example, Mon. гар *(gar)* 'hand' and Tur. *kar-* meaning 'upper arm' are cognates, according to Starling. However, vocabularies do not report 'hand' as a possible meaning for the Turkish word.
- Polymorphism is not considered in cases in which there is a specific word in a language for a meaning, and then one finds in the vocabulary an additional word which means many other things, unless the word is listed as the first translation in the second case. An example is the word угсаа *(ugsaa)* in Mongolian, which according to the vocabulary can have the same meaning of the word цус *(tsus)* 'blood', even though the vocabulary displays 'origin, descent, race, nationality' as its main meaning.
- Polymorphism is not considered in cases in which one word is in a subset/superset relation with respect to another word which is listed for the same meaning. For instance, in the vocabulary one can find Mon. мах *(mah)* for the meaning 'meat', but then one can find 'meat' also under хоол *(xool)*, which is the word for 'food'.

As a consequence, the three polymorphisms mentioned and similar cases have been ignored. Note that this criterion is necessarily obscuring some potential evidence for cognates,

Table 1. The results of Kessler & Lehtonen's (2006) algorithm applied to English, Italian and Hindi.

Pair	Distance	adjusted p-value
English and Italian	3.035	*0.039
English and Hindi	2.997	*0.0268
Hindi and Italian	3.22	*0.0271

but by keeping wordlists short, we are making sure that the true correspondences can reach significant probabilistic levels. In fact, a valid alternative would be to list all the words which are semantically similar (for instance, in the case of 'hand', a potential set would be *arm, biceps, cuff, hand, lower arm, palm, pulse, triceps, upper arm, wrist*) and calculate correspondences over all possible combinations of the word sets. In the case of a set of ten words in two languages for the same meaning, one should take the power set 10 * 10 = 100, and add the 100 pairs to the list of the possible correspondence sets. From the probabilistic viewpoint, it is evident that if, on the one hand, one is probably guaranteed to find several matches, on the other hand significance tests are not necessarily going to provide a positive result, because matches become more likely, and therefore less significant, as the number of comparisons grows. In other words, we are introducing noise that can obscure significant correspondences. This strategy has a further difficulty: one would need an objective criterion to define semantic acceptable sets and make them as wide as their tolerance to semantic shift.

For these reasons, I think that sticking to traditional wordlists and using vocabularies as references is a safer option from the viewpoint of statistical testing.

6 False positives

The methods presented in §3 have already proved to yield some reasonable results when applied to languages known to be related. On the other hand, while Ringe (1992) and Kessler (2001) report cases in which their methods do not yield a positive result when applied to languages which are not provably related, Baxter & Manaster Ramer (2000) and Kessler & Lehtonen (2006) do not present any test case in which their methods correctly return a negative result when two unrelated languages are examined. Therefore, before running the test on Altaic, I ran the algorithms used in this paper on the nine pairs resulting from the combination of the three Altaic languages under investigation and three Indo-European languages: English, Hindi and Italian. The choice of English and Hindi was motivated by the fact that these were the two languages mentioned in Hock & Joseph (1996) and Baxter & Manaster Ramer (2000) as interesting test cases, while Italian was a natural addition given my linguistic competence.

6.1 Testing Kessler & Lehtonen's (2006) algorithm

The first experiment is Kessler & Lehtonen's algorithm applied to the 168-wordlists. The p-value is calculated over 10,000 permutations and the distance between the two lists is reported. In order for the test to succeed, the distance must be significantly smaller than that usually obtained through the permutation tests.

As an initial test, I ran the algorithm on the three Indo-European pairs. The p-values are corrected for multiple comparisons.[7] The results (Table 1) are all positive. According to

[7]When performing multiple comparisons, we might expect some p-values to appear significant just by

Table 2. The results of Kessler & Lehtonen's (2006) algorithm applied to unrelated languages.

Pair	Distance	p-value
Turkish and English	3.574	0.3869
Turkish and Italian	3.924	0.7929
Turkish and Hindi	3.478	0.1333
Mongolian and English	3.688	0.9043
Mongolian and Italian	3.788	0.8347
Mongolian and Hindi	3.47	0.3095
Manchu and English	3.498	0.9044
Manchu and Italian	3.491	0.6699
Manchu and Hindi	3.454	0.4729

Table 3. The results of Baxter & Manaster Ramer's (2000) algorithm applied to English, Italian and Hindi lists.

Pair	Distance	adjusted p-value
English and Italian	0.778	*0.0006
English and Hindi	0.734	*0.0003
Italian and Hindi	0.679	*0.0002

Kessler & Lehtonen's phonetic algorithm, all the pairs are within a .05 significance threshold after the permutation test.

Now, the same experiment was run instead on the spurious pairs, displayed in Table 2. None of the results is significant, and therefore the metric appears indeed robust to false positives.

6.2 Testing Baxter & Manaster Ramer's (2000) algorithm

The same experiment is repeated with the phonological algorithm of Baxter & Manaster Ramer (2000). The test is applied first to the three Indo-European languages.

The results are displayed in Table 3, and they are all highly significant. From these numbers, we may expect the phonological algorithm based on Dolgopolsky's classes to be less conservative with respect to phonetic distances in testing hypotheses of relatedness.

The results found when applying Baxter & Manaster Ramer's algorithm for not provably related pairs are similar to the previous experiment (Table 4). An exception here is represented by Turkish and Hindi, which yield a positive result (p=0.0254), but only before the p-value is corrected for multiple testing: after the adjustment, it becomes not significant. However, there can be a plausible explanation for this pair, because we know independently that the Turkish vocabulary contains many Persian loanwords. Even though words identified as borrowings have been excluded, it is possible that some undetected loanword is increasing the phonetic similarity between the two languages. This problem will be discussed in §8.[8]

chance. The more comparisons we perform, the more the chance of a false positive increases. For this reason, multiplying the p-values for the number of comparisons is needed to make sure that the results are sufficiently robust. The correction chosen is the Holm correction, according to which once the most significant p-value is corrected for the total number of n comparisons by multiplying it for n, it should be removed from the list, so

Table 4. The results of Baxter & Manaster Ramer's (2000) algorithm applied to lists among unrelated languages.

Pair	Distance	p-value	adjusted p-value
Turkish and English	0.887	0.4217	
Turkish and Italian	0.874	0.6065	
Turkish and Hindi	0.797	*0.0254	0.2286
Mongolian and English	0.888	0.4583	
Mongolian and Italian	0.855	0.2263	
Mongolian and Hindi	0.859	0.3584	
Manchu and English	0.922	0.9426	
Manchu and Italian	0.871	0.4029	
Manchu and Hindi	0.869	0.4912	

Table 5. R^2 test on the sound correspondences applied to English, Hindi and Italian.

Pair	R^2	adjusted p-value
English and Italian	178	*0.0021
English and Hindi	106	*0.0014
Italian and Hindi	90	*0.0204

6.3 Testing the presence of sound correspondences

Finally, I ran the R^2 test to look for sound correspondences. The results are shown in Table 5. English and Italian exhibit the highest R^2, and out of 10,000 iterations few higher values are obtained. The situation is similar when comparing English and Hindi. As for Italian and Hindi, the p-value is closer to the threshold, but still significant.

Of course, it was also tempting to apply Ringe's method to the same list of correspondences. The correspondences which are significant, according to the hypergeometric formula, are reported in Table 6.

As we see, there are some differences between the two tests. Ringe's algorithm yields p-values which are closer to the .05 threshold, and in some cases retrieves spurious correspondences (/p/-/k^h/ for English-Hindi, and /p/-/b^h/ and /l/-/tʃ/ for Italian-Hindi). While the case of Italian-Hindi is comparable in the two tests, the tests involving English have a much better result using the R^2 metric. An explanation for this difference is that Ringe's algorithm does not take into account the fact that the /s/ correspondences are very consistent compared to any correspondence found between Italian and Hindi. The property of the R^2 metric of boosting matches which are supported by a high number of correspondence, in this case, is leading the test to yield a positive result for English-Italian and English-Hindi.

The same tests applied to the spurious pairs returned no positive result.

that the second best p-value is corrected for n-1, and so on.

[8] I thank an anonymous reviewer for spotting some words that were not marked as loanwords on the Turkish vocabulary, but were clearly of Persian origin.

Table 6. Number of correspondences among English, Italian and Hindi retrieved through Ringe's (1998) algorithm at the .01 level for the matched lists, with p-values calculated over 10,000 permutations. A value corrected for multiple comparisons is also reported in parentheses.

Eng	Ita	Freq. in Eng	Freq. in Ita	Matches	Distribution	p(n or more)
s	s	30	26	10	*0.0086	
h	k	10	21	5	*0.0041	
n	n	9	9	4	*0.0005	
						0.075
Eng	**Hin**	**Freq. in Eng**	**Freq. in Hin**	**Matches**	**Distribution**	**p(n or more)**
s	s	27	17	8	*0.0044	
n	n	7	6	4	*<0.0001	
m	m	3	6	2	*0.0043	
p	k^h	3	6	2	*0.0043	
						*0.028 (0.056)
Ita	**Hin**	**Freq. in Ita**	**Freq. in Hin**	**Matches**	**Distribution**	**p(n or more)**
d	d	7	7	4	*<0.0001	
n	n	8	6	3	*0.0018	
l	tʃ	9	7	3	*0.0045	
p	b^h	13	2	2	*0.0068	
						*0.024 (0.072)

Table 7. The results of Kessler & Lehtonen's (2006) algorithm applied to Turkish, Manchu and Mongolian.

Pair	Distance	p-value
Turkish and Manchu	3.64	0.4375
Turkish and Mongolian	3.312	0.1013
Mongolian and Manchu	3.584	0.7747

7 Results

7.1 Turkish, Manchu and Mongolian with Kessler & Lehtonen's (2006) algorithm

After the tests of the previous sections, we can now apply the methods to the three Altaic languages.

First, I ran Kessler & Lehtonen's algorithm on the three wordlists. P-values are calculated over 10,000 permutations. None of the results, listed in Table 7, are significant. The closest pair is Turkish and Mongolian, with a distance of 3.312, which is not far from the distance exhibited by Hindi and Italian in §6, but fails the permutation test (p=0.1013). The explanation for this fact is that the phonemic inventories of Turkish and Mongolian resemble each other more than those of Hindi and Italian, and therefore a lower distance is needed in the former case to obtain a positive result.[9]

[9] Turkish-Mongolian randomized lists show an average distance of 3.529, while for Hindi-Italian the average distance is 3.619.

Table 8. The results of Baxter & Manaster Ramer's (2000) algorithm applied to Turkish, Manchu and Mongolian.

Pair	Distance	p-value	adjusted p-value
Turkish and Manchu	0.852	0.4233	
Turkish and Mongolian	0.812	*0.0219	0.0657
Mongolian and Manchu	0.823	0.0703	

Table 9. R^2 test on the sound correspondences applied to Turkish, Manchu and Mongolian.

Pair	R^2	p-value	adjusted p-value
Turkish and Manchu	461	0.1246	
Turkish and Mongolian	584	0.2293	
Mongolian and Manchu	871	0.0011	*0.0033

7.2 Turkish, Manchu and Mongolian with Baxter & Manaster Ramer's (2000) algorithm

Baxter & Manaster Ramer's algorithm is then applied to the same lists. The number of permutations is still 10,000. The results are in Table 8.

The algorithm yields a good result for Turkish-Mongolian (p=0.0219), a non-significant result for Mongolian-Manchu (p=0.0703), and a clear negative result for Turkish-Manchu (p=0.4233). The positive result does not persist after the correction for multiple testing. This global result is interesting, but not totally surprising, given that the two pairs that are close to be significant are the ones for which we may expect areal contact to be present. A similar result was achieved in Oswalt (1998:210-211): even though the wordlists were compiled using evidence from the early stages of the languages and the phonological criteria were different, Oswalt came to the conclusion that the evidence in favor of a Turkic-Tungusic connection was not probative. Given that this is the only pair for which horizontal transmission is unlikely, a test in which this pair fails to show any significant pattern is suspicious, and hints at a possible role of loanwords behind the positive results. §8 will discuss this problem in more detail.

7.3 Evaluating sound correspondences

Finally, I checked the lists for sound correspondences. The results are in Table 9. Here, there is another positive result, this time for Mongolian and Manchu: the Monte Carlo test returns a significant p-value after 10,000 permutations (p=0.0011). In this case, R^2 values are much larger, in absolute terms, than those we have seen in §6: a first explanation for this fact is that given that the Manchu and Mongolian lists contain many polymorphic characters, the number of word pairs evaluated is higher; a second explanation is the fact that the phonemic variation in the Altaic languages is reduced compared to the variation in Indo-European languages: therefore, the correspondences are distributed over a smaller number of possible phonemic matches. The numbers are summarized in Table 10.

This latter point deserves some more attention: if we do consider the phonemic distribution of the languages, we can find many instances of a correspondence just by chance. For instance, the Turkish-Manchu R^2 is supported by thirteen word initial vowel correspondences (/V/-/V/) and nine correspondences between a glide and a vowel (/y/-/V/). Interestingly, if we apply Ringe's hypergeometric test to these two specific pairs (Table 11), none of these

Table 10. Total number of words compared and possible phonemic matches in the languages investigated.

Pair	Words compared	Possible phonemic matches
English and Italian	159	103
English and Hindi	143	103
Italian and Hindi	152	112
Turkish and Manchu	175	84
Turkish and Mongolian	183	79
Mongolian and Manchu	205	100

Table 11. The results of Ringe's (1998) algorithm for /V/-/V/ and /y/-/V/ in Turkish and Manchu.

Tur	Man	Freq. in Tur	Freq. in Man	Matches	Distribution
V	V	35	50	13	0.1425
y	V	24	50	9	0.204

is considered significant, because both /y/ and /V/ are common word-initial phonemes in Turkish. It is precisely because of cases like these ones that one needs a statistical test to tell apart significant correspondences from those predictable given the phonotactics of a pair of languages: an absolute number of correspondences is not probative if we do not control for the initial distribution of the phonemes in the languages.

Now, we can look at the results of Ringe's test applied to every single pair evaluated as significant by the hypergeometric formula (Table 12). The correspondence /k/-/f/ between Turkish and Manchu is not motivated, since none of the word pairs is reported as a cognate in Starling. The same is true for the correspondence between Turkish and Mongolian on /V/-/n/. On the other hand, the correspondences in /d/ have been reported in the Altaic literature as potential cognates. Two entries are presented as cognates in Starling: Tur. *dört*, Mon. дөрөв *(döröv)* 'four', Tur. *duyman* and Mon. дуулах *(duulah)* 'hear'.

As for Manchu and Mongolian, also the correspondence /f/-/V/ has been reported in the Altaic literature: three pairs are listed as cognates in Starling: Mon. үс *(üs)* and Man. *funiyehe* 'hair', Mon. үнс *(üns)* and Man. *fulenggi* 'ash', Mon. улаан *(ulaan)* and Man. *fulgiyan* 'red' (confirmed in Rozycki 1994, but see the doubts in Ligeti 1960). Correspondences in /m/ are also reported: two of them are present in Starling (Mon. могой *(mogoj)* and Man. *meihe* 'snake', Mon. мод *(mod)* and Man. *moo* 'tree'). Apart from the correspondence /V/-/V/, which is difficult to evaluate because it represents a match between two categories rather than a specific sound correspondence, the other two correspondences are not attested.[10]

The correspondences are enough to pass the permutation test, because they yield a p=0.045, which would however be considered not significant if we correct for multiple testing. In the end, this result is very similar to the result we obtained for English and Italian: the hypergeometric test does not take into account the fact that the /f/-/V/ match is more robust

[10] An anonymous reviewer correctly pointed out that /V/-/V/ matches could be the reason why the R^2 yields a test statistic which is more significant than Ringe's algorithm, since the metric gives more weight to matches which are frequent. Since the category /V/ is broad, the result might be biased by general patterns like the presence of vowel infixes or unstressed initial syllables. This is the reason why a match in word-initial vowels is less probative than other sound correspondences.

Table 12. The results of Ringe's (1998) algorithm for Turkish, Manchu and Mongolian with p-value calculated over 10,000 permutations. A value corrected for multiple comparisons is also reported in parentheses.

Tur	Man	Freq. in Tur	Freq. in Man	Matches	Distribution	p(n or more)
k	f	28	15	9	*<0.0001	
						0.628
Tur	Mon	Freq. in Tur	Freq. in Mon	Matches	Distribution	p(n or more)
V	n	37	16	9	*0.0009	
d	d	19	8	4	*0.0045	
						0.304
Man	Mon	Freq. in Man	Freq. in Mon	Matches	Distribution	p(n or more)
V	V	51	57	24	*0.0005	
f	V	15	57	12	*<0.0001	
V	n	51	17	9	*0.0089	
d	x	11	27	5	*0.0073	
m	m	10	10	3	*0.0085	
						*0.045 (0.135)

than any other correspondence that, by chance, could clear the .01 threshold. This is the reason why, following the R^2 result, we can consider this as evidence for a connection between Manchu and Mongolian.

The last question is to what extent these results might have been influenced by areal contact.

8 Loanwords

As already mentioned, it would be interesting to determine how many loanwords are sufficient to cause the methods to yield results which are Type I errors (false positives), i.e., the languages look similar just by virtue of their loanwords. Therefore, I ran an experiment to estimate this threshold for the methods by simulating areal contact.

The experiment is similar to the one performed in Barbançon et al. (2013). The idea in that paper was to test phylogenetic methods by simulating homoplasy and borrowings on artificial lists, to test how the methods were robust to such phenomena. The procedure I adopted is the same. Starting from the list of one among Turkish, Mongolian and Manchu, I created a new list by randomizing the order of the meanings of the starting list. The new list has the same phonemic distribution of the starting one, but since the order of the words is shuffled, it should exhibit no meaningful similarity or correspondence with the starting list once the entries are matched. Then, I forced some lexical borrowing between the lists using different rates of borrowing, in the set {0, 0.05, 0.10}. This means that the two lists are compared in absence of loanwords, in presence of 5% of loanwords, and in presence of 10% of loanwords. To compare the lists, I ran the four algorithms employed in the paper using the standard permutation test. Each experiment is repeated 50 times for each algorithm. At each trial, I kept track of the ratio of loanwords that was sufficient to have two completely randomized lists yielding a significant p-value. The results are in Table 13.

First, we see that with a borrowing rate of 0, some false positives were still identified by all methods, in a range of 0 to 3 out of 50 trials. This is true for all methods, and it means that, contrary to what I showed in the previous section, the methods can yield some false positive even in total absence of borrowing.

The results for actual borrowing simulations are interesting because they show that Baxter & Manaster Ramer's algorithm is clearly the most sensitive to loanwords. In the presence of about 10% of loanwords, the algorithm is almost guaranteed to return a false positive in all cases. If the borrowing rate goes down to 5%, we can still expect to see a false positive almost half of the time. On the contrary, Ringe's algorithm is the most resistant to borrowing, with the other two algorithms in between.

This test proves that the criterion of phonological matching is too generous. Differently from Kessler & Lehtonen's algorithm, it simply assigns a '0' or '1' to a match, with the effect that 10% of the matches are sufficient to return a positive result no matter what happens in the other 90% of the list. On the contrary, by weighting mismatches according to their phonetic distance, Kessler and Lehtonen's test is slightly more conservative.

For this reason, it is clear that in order to accept a successful result for Altaic, a positive result by Baxter & Manaster Ramer's algorithm only would be particularly weak. In general this result raises skepticism towards other applications of the method. In particular, Kassian et al.'s (2015) attempt to argue in favor of Indo-Uralic using Baxter & Ramer's algorithm only becomes weaker. On the other hand, while the other methods are still sensitive to borrowings, their tolerance to loanwords is higher.

Table 13. The results of the experiment on areal contact. Each setting was tested for 50 runs.

Language	Rate of borrowing	K&L	B&MR	R^2	Ringe
Turkish	0	3	3	1	2
Turkish	0.05	16	22	7	3
Turkish	0.10	32	44	30	6
Mongolian	0	2	1	3	1
Mongolian	0.05	9	23	9	1
Mongolian	0.10	33	49	19	10
Manchu	0	3	2	2	1
Manchu	0.05	15	18	8	3
Manchu	0.10	35	48	25	4

9 Conclusion

In this paper, I have addressed the question of the genetic relatedness between Turkish, Mongolian and Manchu using different significance tests that have been proposed in the literature. The tests were run on three modern languages representing the three families traditionally proposed as 'core' Altaic (Turkic, Mongolic and Tungusic). The results are as follows:

- The three Altaic languages were all distinguishable from three Indo-European languages tested here according to all methods. This is an important result, because if these tests were all positive, then we would be facing the problem raised in Comrie (1981) of not being able to draw apart the Altaic signal from a general Eurasiatic signal.

- While Kessler & Lehtonen's method does not yield any positive result, Baxter & Manaster Ramer's algorithm returns a positive match for Turkish-Mongolian, although the result is not significant once the p-values are corrected for multiple comparison.

- As for sound correspondences, the R^2 detects a correspondence pattern between Mongolian and Manchu. Ringe's algorithm retrieves two reported correspondences for Manchu and Mongolian (/f/-/V/) and (/m/-/m/).

- We investigated to what extent the previous results could have been influenced by loanwords. Using Baxter & Manaster Ramer's algorithm, we showed that it is possible to get positive results for a significance test just because of a mild presence of loanwords. While all algorithms suffer from this problem, the thresholds at which this happens were particularly low for this algorithm, and therefore any attempt to use it to show that there are patterns of phonological similarities should be extremely cautious in dealing with potential loanwords.

The results and the methods discussed in this paper confirm that it is possible to design and employ significance tests for long-range comparison.

The aim of this paper was to look for a statistical argument in support of Altaic. At least according to this particular test, it looks like a statistical argument in favor of the Altaic family could not be made, based on the evidence from a restricted list of words belonging to modern languages. Among all results, the fact that the R^2 yields a positive result for Mongolian and Manchu corroborates the hypothesis of their relatedness. Evidence in favor of Turkish and Mongolian only came from Baxter & Ramer's algorithm, even though the positive result disappeared once the p-values were corrected for multiple comparison, and here horizontal transmission cannot be ruled out as an explanation. Finally, no support in favor of a link between Turkish and Manchu, which would be the crucial piece of evidence for Altaic, was detected.

Even though this preliminary test failed, we cannot, of course, interpret this as evidence against the Altaic hypothesis: significance tests are not meant to evaluate how likely a hypothesis is, but are only built to reject alternative hypotheses that can be modeled and controlled through experimental design. However, an argument for Altaic will ultimately require the correspondences to pass a statistical test. As was discussed in this paper, the presence of loanwords and the phonemic distributions of the languages increase the chance of detecting patterns which look very promising, but are not statistically supported once these parameters are properly taken into account mathematically.

Some improvements could make the test more informative. For example, replacing the wordlists with some of the attested or reconstructed ancestral forms is almost guaranteed to yield more encouraging results. The challenges will be: i) to demonstrate that the choice of the forms is unbiased with respect to the hypothesis (and therefore well supported by internal reconstruction); ii) to find an explicit criterion to deal with semantic shift; iii) to defend the results in case they do not pass all the tests. In particular, we have seen that a simple positive test using Baxter & Manaster Ramer's method would be inconclusive, because of its extreme sensitivity to loanwords.

These observations suggest that the 'Altaic controversy' will probably still be under scientific debate in the future.

Résumé Depuis des décennies, des spécialistes en linguistique historique ne s'entendent pas sur la question suivante : la méthode comparative classique fournit-elle assez de preuves pour que l'on considère que les langues altaïques constituent une famille linguistique, comme les familles indo-européenne et ouralienne, ou si la solidité statistique implicite qui sous-tend les correspondances phonologiques régulières fait défaut dans le cas des langues altaïques. Dans cet article, je mène un test statistique sur les listes Swadesh représentant le turc, le mongol et le mandchou pour voir s'il existe des correspondances ou similitudes phonétiques systématiques entre les phonèmes initiaux des mots du vocabulaire de base qui ne sauraient s'expliquer par le hasard. La méthodologie employée s'appuie sur Oswalt (1970), Ringe (1992), Baxter & Manaster Ramer (2000) et Kessler (2001, 2007). Les tests n'indiquent que partiellement l'appartenance à une famille altaïque: le mongol et le mandchou présentent des correspondances phonologiques significatives, tandis que le turc et le mongol présentent une similitude phonétique importante, qui pourrait toutefois être la conséquence d'un contact territorial. Le plus important, c'est que les tests ne montrent aucun rapport significatif entre le turc et le mandchou, quelle que soit la condition.

Zusammanfassung Sprachhistoriker debattieren seit Jahrzehnten darüber, ob die klassische historisch-vergleichende Methode hinreichende Beweise dafür bietet, die altaischen Sprachen als Teil einer einzigen genetischen Einheit wie Indoeuropäisch und Uralisch zu betrachten, oder ob die implizite statistische Robustheit regelmäßiger Lautentsprechungen im Falle des Altaischen fehlt. In diesem Artikel führe ich Signifikanztests mit Swadesh-Listen für Türkisch, Mongolisch und Manchu durch, um festzustellen, ob im Grundwortschatz regelmäßige phonetische Ähnlichkeiten oder Entsprechungen zwischen wortanlautenden Phonemen vorliegen, die nicht als zufällig entstanden angenommen werden können. Die Methodologie stützt sich auf Oswalt (1970), Ringe (1992), Baxter & Manaster Ramer (2000) und Kessler (2001, 2007). Die Tests deuten nur teilweise auf eine altaische Familie hin: Mongolisch und Mandschu weisen signifikante Lautentsprechungen auf. Türkisch und Mongolisch dagegen zeigen zwar einige bedeutende phonetische Ähnlichkeiten, die jedoch auf räumlichen Kontakt zurückzuführen sein könnten. Entscheidend ist, dass die Tests Türkisch und Mandschu unter keinen Bedingungen als verwandt werten.

Author's address
Andrea Ceolin
Department of Linguistics
3401-C Walnut Street
19104 Philadelphia, PA
United States
ceolin@sas.upenn.edu

Acknowledgements Thanks to two anonymous reviewers for detailed comments on the methodology, data and text. Many thanks to Don Ringe, Robin Clark and Mark Liberman for comments on and discussions of this work and other topics related to long-range comparison. Thanks to Ava Irani for the help with Hindi data and Faruk Akkuş for Turkish. Also, thanks to Akiva Bacovcin and David Wilson, the organizers of the Historical Linguistics meeting group at the University of Pennsylvania. All mistakes are mine.

References

Barbançon, François, Steven N. Evans, Luay Nakhleh, Donald A. Ringe & Tandy Warnow. 2013. An experimental study comparing linguistic phylogenetic reconstruction methods. *Diachronica* 30(2). 143–170.

Baxter, William H. 1998. Response to Oswalt and Ringe. In Joseph Salmons & Brian Joseph (eds.), *Nostratic: Sifting the evidence*, 217–236. Amsterdam: John Benjamins.

Baxter, William H. & Alexis Manaster Ramer. 2006. Beyond lumping and splitting: Probabilistic issues in historical linguistics. In Colin Renfrew, April McMahon & Larry Trask (eds.), *Time depth in historical linguistics*, vol. 1, 167–188. Cambridge, England: McDonald Institute for Archaeological Research.

Bomhard, Allan R. 1996. *Indo-European and the Nostratic hypothesis.* Charleston: Signum Desktop Publishing.

Bomhard, Allan R. 2008. *Reconstructing Proto-Nostratic: Comparative phonology, morphology and vocabulary.* Leiden: Brill.

Bomhard, Allan R. 2011. *The Nostratic hypothesis in 2011: Trends and issues.* Washington DC: Institute for the Study of Man.

Comrie, Bernard. 1981. *The languages of the Soviet Union.* Cambridge, England: Cambridge University Press.

Comrie, Bernard. 1998. Regular sound correspondences and long-distance genetic comparison. In Joseph Salmons & Brian Joseph (eds.), *Nostratic: Sifting the evidence*, 271–276. Amsterdam: John Benjamins.

Doerfer, Gerhard. 1973. *Lautgesetz und Zufall: Betrachtungen zum Omnicomparatismus.* Innsbrucker Beiträge zur Sprachwissenschaft.

Dolgopolsky, Aaron B. 1986. A probabilistic hypothesis concerning the oldest relationships among the language families in Northern Eurasia. In Vitalij V. Shevoroshkin & Thomas L. Markey (eds.), *Typology, relationship and time*, 27–50. Ann Arbor: Karoma.

Dybo, Anna & George Starostin. 2008. In defence of the comparative method, or the end of the Vovin controversy. *Papers of the Institute of Oriental and Classical Studies* 19.

Georg, Stefan. 1999. Haupt und Glieder der Altaischen Hypothese: die Körperteilbezeichnungen im Türkischen, Mongolischen und Tungusischen. *Ural-Altaische Jahrbücher* 16. 143–182.

Georg, Stefan. 2008. Review article of Martine Robbeets, 2005. Is Japanese related to Korean, Tungusic, Mongolic and Turkic? *Bochumer Jahrbuch zur Ostasienforschung* 32. 247–278.

Greenberg, Joseph H. 1957. *Essays in linguistics.* Chicago: University of Chicago Press.

Hangin, John G., John R. Krueger & Robert G. Service. 1986. *A modern Mongolian-English dictionary.* Indiana University, Research Institute for Inner Asian Studies.

Hock, Hans Henrich & Brian D Joseph. 1996. *Language change, and language relationship. An introduction to historical and comparative linguistics.* Berlin/New York: Mouton de Gruyter.

Illič-Svityč, Vladislav Markovič. 1971. *Opyt sravnenija nostratičeskych jazykov [An attempt to compare Nostratic languages]* . Moscow: Nauka.

Kassian, Alexei, Mikhail Zhivlov & George Starostin. 2015. Proto-Indo-European-Uralic comparison from the probabilistic point of view. *The Journal of Indo-European Studies* 43(3-4). 301–347.

Kessler, Brett. 2001. *The significance of word lists.* Stanford, California: Center for the Study of Language and Information.

Kessler, Brett. 2007. Word similarity metrics and multilateral comparison. In *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, 6–14. Association for Computational Linguistics.

Kessler, Brett. 2015. Response to Kassian et al., 2015, Proto-Indo-European-Uralic comparison from the probabilistic point of view. *Journal of Indo-European Studies* 43(3-4). 357–367.

Kessler, Brett & Annukka Lehtonen. 2006. Multilateral comparison and significance testing of the Indo-Uralic question. In Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, 33–42. Cambridge, England: McDonald Institute for Archaeological Research.

Li, Gertraude Roth. 2000. *Manchu: a textbook for reading documents.* University of Hawaii Press.

Ligeti, Lajos. 1960. Les anciens éléments mongols dans le mandchou. *Acta Orientalia Academiae Scientiarum Hungaricae* 10(3). 231–248.

Longobardi, Giuseppe, Andrea Ceolin, Luca Bortolussi, Cristina Guardiano, Monica Alexandrina Irimia, Dimitris Michelioudakis, Nina Radkevich & Andrea Sgarro. 2016. Mathematical modeling of grammatical diversity supports the historical reality of formal syntax. In *Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics*, 1–4. Universitätsbibliothek Tübingen.

Longobardi, Giuseppe, Cristina Guardiano, Giuseppina Silvestri, Alessio Boattini & Andrea Ceolin. 2013. Toward a syntactic phylogeny of modern Indo-European languages. *Journal of Historical Linguistics* 3(1). 122–152.

Manaster Ramer, Alexis & Paul Sidwell. 1997. The truth about Strahlenberg's classification of the languages of Northeastern Eurasia. *Journal de la Société Finno-Ougrienne* 87. 139–160.

Menges, Karl Heinrich. 1975. *Altajische Studien: II. Japanisch und Altajisch.* Steiner Franz Verlag.

Miller, Roy Andrew. 1971. *Japanese and the other Altaic languages.* University of Chicago Press.

Miller, Roy Andrew. 1996. *Languages and history: Japanese, Korean, and Altaic.* Bangkok: White Orchid Press.

Nichols, Johanna. 1996. The comparative method as heuristic. In Mark Durie & Malcolm Ross (eds.), *The comparative method reviewed: Regularity and irregularity in language change*, 39–71. Oxford University Press.

Norman, Jerry. 1978. *A concise Manchu-English lexicon.* University of Washington.

Oswalt, Robert L. 1970. The detection of remote linguistic relationships. *Computer Studies in the Humanities and Verbal Behavior* 3(3). 117–129.

Oswalt, Robert L. 1998. A probabilistic evaluation of North Eurasiatic Nostratic. In Joseph Salmons & Brian Joseph (eds.), *Nostratic: Sifting the evidence*, 199–216. Amsterdam: John Benjamins.

Poppe, Nicholas. 1960. *Vergleichende Grammatik Der Altaischen Sprachen; Teil 1: Vergleichende Lautlehre.* Wiesbaden: Harrassowitz.

Poppe, Nicholas. 1965. *Introduction to Altaic linguistics.* Wiesbaden: Harrassowitz.

Ramstedt, Gustav John. 1957. *Introduction to Altaic linguistics.* Moscow: Publishing House of Foreign. lit.

Redhouse, James. 1968. *New Redhouse Turkish-English dictionary.* Publications Department of the American Board.

Ringe, Donald A. 1992. On calculating the factor of chance in language comparison. *Transactions of the American Philosophical Society* 82(1). 1–110.

Ringe, Donald A. 1998. Probabilistic evidence for Indo-Uralic. In Joseph Salmons & Brian Joseph (eds.), *Nostratic: Sifting the evidence*, 153–197. Amsterdam: John Benjamins.

Ringe, Donald A. 2015. Response to Kassian et al., 2015, Proto-Indo-European-Uralic comparison from the probabilistic point of view. *Journal of Indo-European Studies* 43(3-4). 348–356.

Robbeets, Martine. 2005. *Is Japanese Related to Korean, Tungusic, Mongolic and Turkic?* Wiesbaden: Harrassowitz.

Robbeets, Martine. 2015. *Diachrony of verb morphology: Japanese and the Transeurasian languages.* Berlin: De Gruyter Mouton.

Ross, Alan S.C. 1950. Philological probability problems. *Journal of the Royal Statistical Society. Series B (Methodological)* 12.1. 19–59.

Rozycki, William. 1994. *Mongol elements in Manchu.* Bloomington: Indiana University Research Institute for Inner Asian Studies.

Salmons, Joseph & Brian Joseph. 1998. *Nostratic: sifting the evidence.* Amsterdam: John Benjamins.

Sinor, Denis. 1988. *The Uralic languages. Description, history and foreign influences.* Leiden: Brill.

Starostin, Sergei. 1991. On the hypothesis of a genetic connection between the Sino-Tibetan languages and the Yeniseian and North Caucasian languages. In Vitalij V Shevoroshkin (ed.), *Dene-Sino-Caucasian Languages*, 12–41. Ann Arbor: Brockmeyer.

Starostin, Sergei, Anna Dybo, Oleg Mudrak & Ilya Gruntov. 2003. *Etymological dictionary of the Altaic languages.* Leiden: Brill.

Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21(2). 121–137.

Swadesh, Morris. 1971. *The origin and diversification of language.* Piscataway, New Jersey: Transaction Publishers.

Unger, Marshall J. 1990. Summary report of the Altaic panel. *Trends in Linguistics, Studies and Monographs* 45. 479–482.

Villemin, François. 1983. Un essai de détection des origines du japonais à partir de deux méthodes statistiques. In Barron Brainerd (ed.), *Historical linguistics*, 116–135. Bochum: Studienverlag Dr. N. Brockmeyer.

Vovin, Alexander. 2005. The end of the Altaic controversy. In memory of Gerhard Doerfer. *Central Asiatic Journal* 49(1). 71–132.

APPENDIX - Wordlists

NA stands for missing entries, detected borrowings, or compounds.

Meaning	Turkish	Mongolian	Manchu
1 other	NA(1)	бус, өөр (bus, öör)	gūwa
2 one	bir	нэг (neg)	emu
3 two	iki	хоёр (hojor)	juwe
4 three	üç	гурав (gurav)	ilan
5 four	dört	дөрөв (döröv)	duin
6 five	beş	тав (tav)	sunja
7 big	büyük	их, том (ih, tom)	amba
8 long	uzun	урт (urt)	golmin
9 wide	geniş	өргөн (örgön)	onco, leli
10 thick	kalın	өтгөн (ötgön)	jiramin, fisin
11 heavy	ağır	хүнд (hünd)	ujen
12 small	küçük	жижиг, жаал (žižig, žaal)	ajige
13 short	kısa	ахар, богино (ahar, bogino)	foholon, fangkala
14 narrow	dar	нарийхан, ухал (narijhan, uhal)	isheliyen
15 thin	ince	шингэн , нимгэн (šingen , nimgen)	nekeliyen, narhun
16 man	er, erkek	эр (er)	NA(2)
17 person	kişi	хүн (hün)	niyalma
18 child	çocuk	хүүхэд (hüühed)	jui
19 animal	NA(3)	амьтан (amtan)	ergengge
20 fish	balık	загас (zagas)	nimaha
21 bird	kuş	шувуу (šuvuu)	gasha, cecike
22 dog	köpek, it	нохой (nohoj)	indahūn
23 louse	bit	бөөс (böös)	cihe
24 snake	yılan	могой (mogoj)	meihe
25 worm	kurt	өт, хорхой (öt, horhoj)	umiyaha, beten
26 tree	ağaç	мод (mod)	moo
27 forest	orman	ой (oj)	bujan, weji
28 fruit	yemiş	жимс (žims)	tubihe
29 seed	NA(4)	үр (ür)	use
30 leaf	yaprak	навч (navč)	abdaha
31 root	kök	уг, үндэс (ug , ündes)	fulehe, da
32 bark	kabuk	холтос (holtos)	notho
33 flower	çiçek	цэцэг (tsetseg)	ilha
34 grass	ot	өвс, ногоо (övs, nogoo)	orho, niyanciha
35 rope	ip	олс (ols)	futa
36 skin	deri	арьс (ars)	sukū
37 meat	et	мах (mah)	yali
38 blood	kan	цус (tsus)	senggi
39 bone	kemik	яс (jas)	giranggi
40 fat	yağ	өөх (ööh)	tarhūn
41 egg	yumurta	өндөг (öndög)	umhan
42 horn	boynuz	эвэр, бүрээ (ever, büree)	uihe

Meaning	Turkish	Mongolian	Manchu
43 tail	kuyruk	сүүл (süül)	uncehen
44 feather	tüy	өд (öd)	funggaha
45 hair	kıl, saç	үс (üs)	funiyehe
46 head	baş	толгой (tolgoj)	uju
47 ear	kulak	чих (čih)	šan
48 eye	göz	нүд (nüd)	yasa
49 nose	burun	хамар (hamar)	oforo
50 mouth	ağız	ам (am)	angga
51 tooth	diş	шүд (šüd)	weihe
52 tongue	dil	хэл (hel)	ilenggu
53 fingernail	tırnak	хумс (hums)	hitahūn
54 foot	ayak	хөл (höl)	bethe
55 knee	diz	өвдөг (övdög)	tobgiya
56 hand	el	гар (gar)	gala
57 wing	kanat	далавч (dalavč)	asha
58 belly	karın	гэдэс (gedes)	NA(5)
59 guts	bağırsak	NA(6)	duha
60 neck	boyun	хүзүү (hüzüü)	monggon
61 back	arka, sırt	нуруу (nuruu)	fisa
62 heart	yürek	зүрх (zürh)	niyaman
63 to drink	içmek	уух (uuh)	omimbi
64 to eat	yemek	идэх (ideh)	jembi
65 to bite	ısırmak	хазах (hazah)	saimbi
66 to spit	tükürmek	нулимах (nulimah)	cifelembi
67 to vomit	kusmak	бөөлжих (böölžih)	jurumbi, fudambi
68 to blow	üflemek	үлээх (üleeh)	fulgiyembi, edumbi
69 to breathe	soluk almak	амьсгалах (amsgalah)	ergen gaimbi
70 to laugh	gülmek	инээх (ineeh)	injembi
71 to see	görmek	үзэх, харах (üzeh, harah)	sabumbi
72 to hear	duymak	сонсох, дуулах (sonsoh, duulah)	donjimbi
73 to know	bilmek	мэдэх (medeh)	sambi
74 to think	NA(7)	бодох (bodoh)	gūnimbi
75 to smell	koklamak	үнэрлэх (ünerleh)	wangkiyambi
76 to fear	korkmak	айх (ajh)	gelembi, olhombi
77 to sleep	uyumak	унтах, нойрсох (untah, nojrsoh)	amgambi
78 to live	yaşamak	амьдрах (amdrah)	banjimbi
79 to die	ölmek	үхэх (üheh)	bucembi
80 to fight	savaşmak	байлдах, зодолдох (bajldah, zodoldoh)	afambi
81 to hunt	avlamak	авлах (avlah)	abalambi, tumbi
82 to hit	vurmak	цохих (tsohih)	tantambi
83 to cut	kesmek	огтлох, хэсэглэх (ogtloh, hesegleh)	faitambi, furumbi
84 to split	yarmak	хагалах (hagalah)	delhebumbi, dendebumbi
85 to stab	bıçaklamak	хутгалах (hutgalah)	tokombi

Meaning	Turkish	Mongolian	Manchu
86 to scratch	NA(8)	маажих (maažih)	wašambi
87 to dig	kazmak	малтах, ухах (maltah, uhah)	fetembi
88 to swim	yüzmek	сэлэх (seleh)	ebišembi
89 to fly	uçmak	нисэх (niseh)	deyembi
90 to walk	yürümek	алхах (alhah)	yabumbi
91 to come	gelmek	ирэх (ireh)	jimbi
92 to lie	yatmak	хэвтэх (hevteh)	dedumbi
93 to sit	oturmak	суух (suuh)	tembi
94 to stand	durmak	зогсох (zogsoh)	ilimbi
95 to turn	dönmek	эргэх (ergeh)	forgošombi
96 to fall	düşmek	унах (unah)	tuhembi
97 to give	vermek	өгөх (ögöh)	bumbi
98 to hold	tutmak	барих (barih)	jafambi
99 to squeeze	ezmek	шахах (šahah)	hafitambi
100 to rub	ovmak	үрэх (üreh)	monjimbi
101 to wash	yıkamak	угаах (ugaah)	obombi
102 to wipe	silmek	арчих (arčih)	mabulambi
103 to pull	çekmek	татах (tatah)	tatambi
104 to push	itmek	түлхэх, чихэх (tülheh, čiheh)	anambi
105 to throw	atmak	хаях, шидэх (hayah, šideh)	maktambi
106 to tie	bağlamak	уях, баглах (uyah, baglah)	huthumbi, hūwaitambi
107 to sew	dikmek	оёх (ojoh)	ufimbi
108 to count	saymak	тоолох (tooloh)	tolombi
109 to say	söylemek	хэлэх (heleh)	hendumbi, sembi
110 to sing	sarki söyle	NA(9)	uculembi
111 to play	oynamak	наадах, тоглох (naadah, togloh)	efimbi
112 to float	yüzmek	хөвөх (hovoh)	dekdembi
113 to flow	akmak	урсах (ursah)	eyembi
114 to freeze	donmak	хөлдөх (höldöh)	gecembi
115 to swell	şişmek	хөөх, хавдах (hooh, havdah)	aibimbi, dukdurembi
116 sun	NA(10)	нар (nar)	šun
117 moon	ay	сар (sar)	biya
118 star	yıldız	од (od)	usiha
119 water	su	ус (us)	muke
120 rain	yağmur	бороо (boroo)	aga
121 river	ırmak	гол (gol)	bira
122 lake	göl	нуур (nuur)	omo, tenggin
123 sea	deniz	далай, тэнгис (dalaj, tengis))	mederi, namu
124 salt	tuz	давс (davs)	NA(11)
125 stone	taş	чулуу (čuluu)	wehe
126 sand	kum	элс (els)	yonggan
127 earth	toprak	товрог (tovrog)	na
128 cloud	bulut	үүл (üül)	tugi
129 fog	sis	манан (manan)	talman
130 sky	gök	огторгуй, тэнгэр (ogtorguj, tenger)	abka

Meaning	Turkish	Mongolian	Manchu
131 wind	yel	салхи (salhi)	edun
132 snow	kar	цас (tsas)	nimanggi
133 ice	buz	мөс (mös)	juhe
134 smoke	duman	утаа (utaa)	NA(12)
135 fire	NA(13)	гал (gal)	tuwa
136 ash	kül	үнс (üns)	fulenggi
137 to burn	yan-	шатах (šataah)	deijimbi
138 road	yol	зам (zam)	jugūn
139 mountain	dağ	уул (uul)	alin
140 red	kızıl	улаан (ulaan)	fulgiyan
141 green	yeşil	ногоон (nogoon)	niowanggiyan
142 yellow	sarı	шар (šar)	suwayan
143 white	ak, beyaz	цагаан (tsagaan)	NA(14)
144 black	kara, siyah	хар (har)	sahaliyan
145 night	gece	шөнө (šönö)	dobori
146 day	gün	өдөр (ödör)	inenggi
147 year	yıl	жил (žil)	aniya
148 warm	sıcak	бүлээн, дулаан (büleen, dulaan)	bulukan, halukan
149 cold	soğuk	хүйтэн (hüjten)	šahūrun
150 full	dolu	дүүрэн (düüren)	jalu
151 new	yeni	шинэ (šine)	ice
152 old	eski	хуучин (huučin)	fe
153 good	iyi	сайн (sajn)	sain
154 bad	kötü	муу (muu)	ehe
155 rotten	çürük	ялзахсай (yalzahsaj)	niya-
156 dirty	kirli	бохир, хиртэй (bohir, hirtej)	nantuhūn, langse
157 straight	düz	шулуун (šuluun)	sijirhūn
158 round	yuvarlak	дугуй (duguj)	muheliyen
159 sharp	NA(15)	хурц (hurts)	dacun
160 dull	NA(16)	мохоо (mohoo)	modo
161 smooth	düz	гөлгөр (gölgör)	bišun
162 wet	ıslak	нойтон (nojton)	usihin
163 dry	kuru	хуурай (huuraj)	olhon
164 correct	doğru	зөв (zöv)	tob
165 near	yakın	ойр (ojr)	hanci, hamika
166 right	sağ	баруун (baruun)	ici
167 left	sol	зүүн (züün)	hashū
168 name	ad	нэр (ner)	gebu

Meaning	English	Italian	Hindi
1 other	other	altro	dūsrā
2 one	one	uno	ek
3 two	two	due	do
4 three	three	tree	tīn
5 four	four	quattro	cār
6 five	five	cinque	pā̃c
7 big	big	grande	baṛā
8 long	long	lungo	lambā
9 wide	wide	largo	cauṛā
10 thick	thick	spesso	gāṛhā
11 heavy	heavy	pesante	bhārī
12 small	small	piccolo	choṭā
13 short	short	corto	NA(17)
14 narrow	narrow	stretto	NA(18)
15 thin	thin	sottile	patlā
16 man	man	uomo	NA(19)
17 person	NA(20)	persona	NA(21)
18 child	child	bambino	bacṛhā
19 animal	NA(22)	animale	NA(23)
20 fish	fish	pesce	machlī
21 bird	bird	uccello	ciṛiyā
22 dog	dog	cane	kuttā
23 louse	louse	pidocchio	jũ
24 snake	snake	serpente	sā̃p
25 worm	worm	verme	kīṛā
26 tree	tree	albero	peṛ
27 forest	NA(24)	foresta	jaṅgal
28 fruit	NA(25)	frutta	phal
29 seed	seed	seme	bīj
30 leaf	leaf	foglia	pattā
31 root	root	radice	jaṛ
32 bark	bark	corteccia	chāl
33 flower	NA(26)	fiore	phūl
34 grass	grass	erba	ghās
35 rope	rope	corda	rassī
36 skin	skin	pelle	tvacā
37 meat	meat	carne	mās
38 blood	blood	sangue	NA(27)
39 bone	bone	osso	haḍḍī
40 fat	fat	grasso	carbī
41 egg	egg	uovo	aṇḍā
42 horn	horn	corno	sīṅg

Meaning	English	Italian	Hindi
43 tail	tail	coda	pūnch
44 feather	feather	piuma	paṅkh
45 hair	hair	capelli	bāl
46 head	head	testa	sir
47 ear	ear	orecchio	kān
48 eye	eye	occhio	ā̃kh
49 nose	nose	naso	nāk
50 mouth	mouth	bocca	mū̃h
51 tooth	tooth	dente	dā̃t
52 tongue	tongue	lingua	jībh
53 fingernail	fingernail	unghia	nākhun
54 foot	foot	piede	pair
55 knee	knee	ginocchio	ghuṭnā
56 hand	hand	mano	hāth
57 wing	wing	ala	paṅkh
58 belly	belly	pancia	peṭ
59 guts	guts	intestino	antṛī
60 neck	neck	collo	NA(28)
61 back	back	schiena	pīṭh
62 heart	heart	cuore	hṛ́day
63 to drink	to drink	bere	pīnā
64 to eat	to eat	mangiare	khānā
65 to bite	to bite	mordere	kāṭnā
66 to spit	to spit	sputare	thūknā
67 to vomit	NA(29)	vomitare	ulṭī karnā
68 to blow	to blow	soffiare	phū̃k mārnā
69 to breathe	to breathe	respirare	sā̃s lenā
70 to laugh	to laugh	ridere	hasnā
71 to see	to see	vedere	dekhnā
72 to hear	to hear	sentire	sunnā
73 to know	to know	sapere	jānnā
74 to think	to think	pensare	socnā
75 to smell	to smell	annusare	sūṅghnā
76 to fear	to fear	temere	ḍarnā
77 to sleep	to sleep	dormire	sonā
78 to live	to live	vivere	jīnā
79 to die	to die	morire	NA(30)
80 to fight	to fight	combattere	laṛnā
81 to hunt	to hunt	cacciare	śikār karnā
82 to hit	to hit	colpire	mārnā
83 to cut	to cut	tagliare	kātnā
84 to split	to split	fendere	baṇṭnā
85 to stab	to stab	pugnalare	bhoṅknā

Meaning	English	Italian	Hindi
86 to scratch	to scratch	graffiare	kharoñcnā
87 to dig	to dig	scavare	khodnā
88 to swim	to swim	nuotare	taiṛnā
89 to fly	to fly	volare	uṛnā
90 to walk	to walk	camminare	calnā
91 to come	to come	venire	ānā
92 to lie	to lie	giacere	letnā
93 to sit	to sit	sedere	baiṭhnā
94 to stand	to stand	stare in piedi	khaṛā honā
95 to turn	to turn	girare	muṛnā
96 to fall	to fall	cadere	girnā
97 to give	to give	dare	denā
98 to hold	to hold	tenere	pakaṛnā
99 to squeeze	to squeeze	spremere	ghusā denā
100 to rub	to rub	strofinare	malnā
101 to wash	to wash	lavare	dhonā
102 to wipe	to wipe	asciugare	poñchnā
103 to pull	to pull	tirare	khı̄ñcnā
104 to push	to push	spingere	dhakkā denā
105 to throw	to throw	lanciare	pheṅknā
106 to tie	to tie	legare	bādhnā
107 to sew	to sew	cucire	sı̄nā
108 to count	NA(31)	contare	ginnā
109 to say	to say	dire	kahnā
110 to sing	to sing	cantare	gānā
111 to play	to play	giocare	khelnā
112 to float	to float	galleggiare	taiṛnā
113 to flow	to flow	fluire	bahnā
114 to freeze	to freeze	gelare	jamnā
115 to swell	to swell	gonfiarsi	sūjnā
116 sun	sun	sole	sūraj
117 moon	moon	luna	cānd
118 star	star	stella	tārā
119 water	water	acqua	pānı̄
120 rain	rain	pioggia	varṣā
121 river	river	fiume	nadı̄
122 lake	NA(32)	lago	jhı̄l
123 sea	sea	mare	samandar
124 salt	salt	sale	NA(33)
125 stone	stone	pietra	patthar
126 sand	sand	sabbia	ret
127 earth	earth	terra	dhartı̄
128 cloud	cloud	nuvola	bādal
129 fog	fog	nebbia	dhundh
130 sky	sky	cielo	NA(34)

Meaning	English	Italian	Hindi
131 wind	wind	vento	NA(35)
132 snow	snow	neve	NA(36)
133 ice	ice	ghiaccio	NA(37)
134 smoke	smoke	fumo	dhuā
135 fire	fire	fuoco	āg
136 ash	ash	ceneri	rākh
137 to burn	to burn	bruciare	jalnā
138 road	road	strada	saṛak
139 mountain	NA(38)	montagna	pahāṛ
140 red	red	rosso	lāl
141 green	green	verde	harā
142 yellow	yellow	giallo	pīlā
143 white	white	bianco	NA(39)
144 black	black	nero	kālā
145 night	night	notte	rāt
146 day	day	giorno	din
147 year	year	anno	NA(40)
148 warm	warm	caldo	NA(41)
149 cold	cold	freddo	ṭhaṇḍ
150 full	full	pieno	pūrā
151 new	new	nuovo	nayā
152 old	old	vecchio	purānā
153 good	good	buono	acchā
154 bad	bad	cattivo	burā
155 rotten	rotten	marcio	saṛā
156 dirty	dirty	sporco	gandā
157 straight	straight	dritto	sīdhā
158 round	NA(42)	rotondo	gol
159 sharp	sharp	aguzzo, affilato	tīkhā
160 dull	dull	smussato	kund
161 smooth	smooth	liscio	ciknā
162 wet	wet	bagnato	gīlā
163 dry	dry	asciutto, secco	sūkhā
164 correct	correct	corretto	NA(43)
165 near	near	vicino	nazdīk
166 right	right	destra	dāyā
167 left	left	sinistra	bāyā
168 name	name	nome	nām

(1) diğer, borrowed from Persian
(2) haha, onomatopoeia (cf. Robbeets 2005 for the same entry in Japanese)
(3) hayvan, borrowed from Arabic
(4) tohum, borrowed from Persian
(5) hefeli, borrowing from Mongolian (Rozyzcki 1994)
(6) same meaning of 'belly'
(7) düşünmek, same root as düşmek 'to fall'
(8) tırmalamak, same root as tırnak 'fingernail'
(9) same meaning of 'to hear'
(10) güneş, same root as gün 'day'
(11) dabsun, borrowing from Mongolian (Rozyzcki 1994)
(12) šanggiyan, same root as šanyan 'white'
(13) āteş, borrowed from Persian
(14) šanyan, borrowed from Mongolian (Ligeti 1960)
(15) keskin, same root as kesmek 'to cut'
(16) kör, borrowed from Persian
(17) same meaning of 'small'
(18) tang, borrowed from Persian
(19) ādmī, borrowed from Persian
(20) person, borrowed from French
(21) insān, borrowed from Persian
(22) animal, borrowed from French
(23) jānvar, borrowed from Persian
(24) forest, borrowed from Old French
(25) fruit, borrowed from Old French
(26) flower, borrowed from French
(27) xun, borrowed from Persian
(28) gardan, borrowed from Persian
(29) to vomit, borrowed from Latin
(30) same meaning of 'hit'
(31) to count, borrowed from French
(32) lake, borrowed from Old French
(33) namak, borrowed from Persian
(34) āsmān, borrowed from Persian
(35) havā, borrowed from Persian
(36) barat, borrowed from Persian
(37) same meaning of 'snow'
(38) mountain, borrowed from French
(39) safed, borrowed from Persian
(40) sāl, borrowed from Persian
(41) garm, borrowed from Persian
(42) round, borrowed from Old French
(43) sahī, borrowed from Persian