S. A. Burlak
Institute of Oriental Studies, Moscow

Languages, DNA, relationship and contacts

In this paper, language contacts are classified according to their results that can be observed by means of historical and comparative linguistics. Various types of bilingual contacts and of language shift are discriminated; they differ in the way they affect vocabulary, grammar and phonetics. These differences are connected with the demographic situation; thus, looking at a language that underwent contact influence, one can say what type of contact could have produced such results. Such information about prehistoric communities can help to reconcile linguistic evidence with archaeological and genetic data in order to produce a more detailed picture of the history of peoples and their languages.

Keywords: language relationship, language contacts, language shift.

In the beginning of the 20th century, Gustaf Kossinna (Kossinna 1911) put forward the hypothesis that material culture correlates with language and ethnicity, which is now known to be wrong (see e.g., Kuz'menko 2011). Ethnic identity need not fully correlate with either language or with genetic features, although there are, indeed, numerous cases of such a coincidence. There have been numerous attempts to reconcile linguistic data with archeological and genetical evidence (see especially Blench & Spriggs 1997, 1998, 1999a, 199b): e.g., Yu. Kuz'menko in his recent book about early Germans considers Werner's law in Proto-Germanic as one of the traces of contacts between ancient Germans and Finno-Ugrians, corresponding to archaeologically documented contacts between Neolithic cultures of Northern Europe and the Pit-Comb Ware culture and genetically documented peoples having Indo-European haplogroup R1a and Uralic haplogroup N (Kuz'menko 2011). But in many cases it is not so easy to coordinate linguistic, genetic and archaeological data. P. Forster and C. Renfrew (2011) argue that Y-chromosome is a good predictor of language relationship, but there are some deviating cases, e.g., "in Greenland, where both the mtDNA … and the language today are nearly pure Eskimo, while half the Y chromosomes are European … evidently from contact with male European whalers over the centuries" (Forster, Renfrew 2011: 1391 with further references). There are historically known migrations that remain invisible to archaeological and/or genetic methods — the Dorian invasion, thought to be the cause of destruction of the Mycenaean culture, is not traced archaeologically until the Xth century BC; Hungarians, being migrants from the Trans-Ural region, do not differ genetically from the neighboring peoples of Central Europe (Balanovsky et al. 2013); Saami, despite having parallels with Siberian peoples in mythology (Beryozkin 2008), dermatoglyphics (Khit' 1983: 81; 1991: 65) and some cultural features (Gurvich 1968, Simchenko 1976: 255–261), show almost no genetic relations to Siberian populations: there is only one "Saami-Siberian" genetic marker, mitochondrial haplogroup Z, and its frequency in Saami is diminishing — only about 1% remains (Balanovsky 2008); genetic data contradict the hypothesis (based on linguistic data) of prehistoric Indo-European migrations into Hindustan (Sengupta & al. 2006, Sahoo & al. 2006, Sharma & al. 2009, Underhill 2009, Metspalu & al. 2011). I. Peiros supposes that "a direct correlation between [the linguistic and archaeological] accounts is theoretically impossible and the sole link between

them is the community itself" (Peiros 1997: 155). Thus, the question arises whether comparative linguistics can recover additional information about prehistoric communities.

Comparative linguistics produces genealogical trees, but a pure "tree-like" structure of languages' history can be observed only when a group of people comes to uninhabited territories and nobody else comes there after them — and this does not happen too often. Thus, the main part of languages' history is the history of language contacts (reflecting ethnic and cultural contacts). When linguists construct genealogical trees of language families, they sometimes imagine language contact as a short process during which a language obtains several loanwords from another language; and the main task is to separate them off from indigenous words to establish true cognates. But language contacts may be of various types, and accordingly produce different results that can be noticed by comparative linguistics, genetics, archaeology and comparative mythology, with the difference depending on the type of contact.

Literature on language contact is abundant (see, e.g., Thomason & Kaufman 1988, Sankoff 2001, Trudgill 2011). This research mainly deals with grammatical change. Discussion topics include such issues as whether grammar reveals genetic relationship better than vocabulary (it does not; see Thomason & Kaufman 1988), and whether, as a result of contacts, it becomes more simple (Trudgill 2004; McWhorter 2005), or more complex (Nichols 1992), or neither (because "simplification at one structure point is likely to be offset by complication at another" Thomason 2008: 3). In this paper, I will not discuss these topics; I am going to concentrate upon those aspects of language contacts that can be observed by historical and comparative linguistics and classify language contacts according to their results in surviving languages, genes, myths and archaeological cultures. "Language U" (and, respectively, "people U", "culture U") will designate the language that undergoes contact influence but survives; "language I" (and, respectively, "people I", "culture I") is the language that is the source of influence.

The first type of language contact can be called a "shallow" contact. It is the type that has already been described — leading to several loanwords in language U, and nothing more.

A subtype of it may be called a "xenophobic" contact: speakers of the language U do not know and do not want to know the language I, considering language U to be the only "authentic" human language; it follows that all loanwords from language I must be interpreted in terms of language U. For example, Russian *верстак* 'workbench' is a loanword from German *Werkstatt*; in German, it consists of *Werk* 'work' and *Statt* 'place', but in Russian, it is reanalyzed as if it were derived from the word *верста* 'verst (a measure of distance)' with the suffix *-ak*. Similarly, Eng. *woodchuck* 'a marmot, *Marmota monax*' does not chuck wood — this word comes from Cree (Algonquian) *otchek* or Ojibwa *otchig* 'marten'. There are lots of such loanwords in Chinese, e.g., 英國 *yīng guó* 'England', lit. "a heroic land".

This type of contact appears when the speakers of language U "feel negatively about bilingualism or are too attached to monolingualism" (Cook 2008: 138).

When languages undergo such contacts, no grammar interference occurs, and basic vocabulary is not affected: Russian has no traces of German grammar, and there are no words of German origin in the Russian Swadesh 100-word list. Neither does English show any traces of Algonquian grammar, and not a single word in the English Swadesh 100-word list comes from Algonquian.

In such cases, loanwords are restricted to certain technology, local environmental or ethnic peculiarities. In material culture, there will be some imports from I, or some technologies borrowed from the culture I. For example, Russian joinery has borrowed many tools (as well as many terms) from German.

If a language (and a people, respectively) undergoes a contact of this type in its history, the linguist will have to deal with several examples of folk etymology (violating regularity in

phonetic correspondences), a mythologist and a geneticist may see some minor admixture of I material to U material (or no admixture at all), and an archaeologist will find some imports from I culture to U culture.

The next group of types involves bilingualism. A considerable part of the U people (maybe all of them) are bilingual, knowing both U and I languages, but in their everyday life they are able to make extensive use of U. In this case, language U retains its basic vocabulary but acquires many lexical borrowings in the cultural sphere. When such a contact situation endures for a long time period, a mechanism is formed that makes it possible to incorporate any language I word into language U: each phoneme of language I receives its own substitute in language U, thus forming a system of regular phonetic correspondences. If the languages are closely related, there will be a higher percentage of mutually retained archaisms, and examples of "phoneme-to-phoneme conversion" may occur (see Burlak, Starostin 2005: 63).

Such might have been the case with the word for 'king' in Slavic languages: it was borrowed from Germ. *Karl* (the name of Charlemagne, 742–814) and yielded Bulg. *крáлят*, Serb. *крâљ*, Pol. *król*, Rus. *король*, as if it were a Slavic word of Proto-Indo-European origin. Sometimes, words are only partly converted, causing irregularities in phonetic correspondences. E.g., the Tocharian A word *warpiśke* 'garden' is borrowed from Tocharian B *werpiśke* 'id.'; the root vowel is converted (A *a* ~ B *e* is a phonetic correspondence of great frequency), whereas the suffix is borrowed in full, so that even the final *e* (which is always dropped in original words of Tocharian A) is preserved. When the languages under contact are closely related, loan-translations (showing the same regular phonetic correspondences as the inherited words) can occur. An example of this can be seen in Siouan languages: the name for distilled spirits is Ponca *ppé:de-ni*, Kansa *ppé:ǰe-ni*, Osage *hpé:te-ni* etc., as if it were inherited from Proto-Siouan 'fire' + 'water', though "of course the Siouan-speaking peoples did not have distilled liquor until post-contact times, and the compound came about either through parallel innovation, based on the properties of the liquid, or through contact with Algonquian-speaking peoples to the east who had a similar compound (equally non-reconstructible) from which the Siouan could easily have been loan-translated. It could even represent a back-translation by whites of the Algonquian pattern" (Rankin 2003: 193).

There are two types of bilingual situation, depending on whether there are mixed settlements or not. When there are no mixed settlements, the U people are able to use language U in their everyday routine, their children are introduced to the standard variant of U, and I is acquired as a second language. But competence in the second language affects the skills of speaking the first one (Cook et al., 2007). As V. Cook notes, "French and Spanish learners of English have their voice onset time affected by their knowledge of English, so that to some extent they have a single system they use in both languages. English speakers of Japanese use aizuchi (nodding for agreement) when talking English (Locastro, 1987). Experiments with syntax have shown unexpected effects on the first language from knowing a second language. Hartsuiker et al. (2004) found, for instance, that hearing passives in one language increased their production in using another" (Cook 2008: 232). Therefore, not only lexical, but also some phonetic, semantic and grammatical borrowings may penetrate into language U, although they would never be numerous and profound. For example, many Russians know English and use many words of English origin: *джинсы* (< *jeans*), *компьютер* (< *computer*), *девайс* (< *device*), *эксклюзивный* (< *exclusive*) and so on, but none of these bilingual speakers tend to replace basic Russian words with their English equivalents. In grammar, minor interference can be observed: there are some derivatives with the suffix -*ing*, e.g., *интернет-ПИФинг* from *ПИФ* (an abbreviation of *паевой инвестиционный фонд* 'unit investment fund'), and sometimes constructions with a prepositional noun attribute, not marked for case (instead of using an adjec-

tive or a noun in Genitive case in postposition, see Levontina 2010: 208–216) are used. In Veps, a language in contact with Russian, the verb meaning 'to separate from smb.' began to be used with Comitative case (and not with Ablative or the like) as a result of influence on the part of Russian *расстаться с кем-л.* 'to separate from smb.' (Rus. *с* means 'with', and not 'from').

In such a situation, the material culture will be mixed, since all the technological innovations of I will have spread to U, but the people U will retain some peculiarities (headdresses, ornaments, mortuary rites and so on) of symbolic rather than utilitarian nature (cf. "deep" and "surface" representations in the terminology of I. Peiros, see Peiros 1997: 156). For example, English-speaking Russians use computers and wear jeans, but are buried according to the customs of the Russian Orthodox church. Archi people, speaking the Archi language (Lezgic group) and contacting with the Avar language (Avar-Andic group), have borrowed many elements of dress from Avar people, but retained the *chuhta* — an aboriginal women's headdress.

The geneticist, in such a situation, may see a minimal (or even zero) admixture of I; likewise, the myths will also mostly belong to U.

In the situation of mixed settlements, the contact situation changes considerably. Probably, the only possibility for a non-dominant language to survive under these circumstances is to be claimed an important marker of ethnic identity and to be valued accordingly. For example, in Kupwar (a village in the Sangli district, Maharashtra, India) Kannada language (Dravidian family), Urdu and Marathi (Indo-Aryan group of the Indo-European family) are all spoken, and such a multilingual situation has already lasted for several centuries (Gumperz & Wilson 1971). J. J. Gumperz and R. Wilson have studied this case at length and found out that traces of grammar interference in all the three languages are abundant: word order principles, semantics of gender, means of expression of indirect speech, clusivity, agreement and many other elements of grammar have converged to a great extent; in many cases, it is possible to translate a sentence from one language into another "by simple morph for morph substitution" (Gumperz & Wilson 1971: 155), and even a few grammar markers are borrowed. At the same time, lexical borrowings, though numerous enough, are almost fully restricted to the cultural sphere; basic vocabulary is not affected. The sociolinguistic setting underlying this language contact situation is the following (Gumperz & Wilson 1971): at home, people speak only one home-group language, and each language group (being also a culture group) "has its own neighborhood" (Gumperz & Wilson 1971: 153). Thus, language is a marker of the community, and to address the speaker of a different language in one's own language "would be tantamount to saying I consider you a potential member of my home and friendship group" (Gumperz & Wilson 1971: 153). Naturally, each group has its own religion (Kannada speakers are Jains, Urdu speakers are Moslems and Marathi speakers are untouchables), and religious rituals are conducted in the language of the corresponding group; dress is also one of the group markers.

In this case, each language is undergoing influence (language U) and at the same time is the source of influence (language I) for the other languages.

As a result of such contact, the linguist will find a language with inherited basic vocabulary, lots of borrowings in the cultural sphere (showing regular phonetic correspondences) and a grammar that cannot be traced back to the protolanguage. The geneticist and the comparative mythologist will find some admixture of I material to U data. The archaeologist will find a settlement with different types of mortuary rites in the same chronological horizon. Such may have been the case in Khaznidon (Northern Ossetia), where there are burials of two types: in the edgewise position (kurgan 2, burial 1) and in supine position (kurgan 5, burials 2 and 4, see Nikolaeva 2011: 407, 414–416).

When a community uses language as an ethnic marker, even closely related languages can survive. If the community practises exogamy, there may be gender differences in genes and material culture. A good example of such a situation is the contact between the Waimaha (Bará) and Tuyuka languages of the Tucano family (Vaupes river, South America), studied by Jean Jackson (Jackson 1974). These Indians can marry only those whose native language is not the same as their own. Thus, they endeavor to keep their language pure: "I would shame the longhouse if I learned to speak Bará with Tuyuka words", said one of the Bará women to Jean Jackson (Jackson 1974: 63). In such a longhouse there are about 20 families, all the men are of the same ethnic group and the same language, while all the women are from different ethnic groups and may speak different languages. The Tucano language serves as a *lingua franca*. Myths have some motifs in common, for example, L22 according to Yu. Beryozkin's catalogue (Beryozkin n.d.): an evil spirit comes to people that are sleeping and takes their eyes out; then blinded people turn into monkeys (in the traditions of the other neighboring ethnic groups, blinded people do not turn into monkeys).

The next group of types consists of various kinds of language shifts.

The "acculturation model" (Cook 2008, Gardner 2007) depends to a great extent on language attitude, namely, whether it is considered to be of value to maintain (a) "cultural identity and characteristics" (Cook 2008: 141) including language and (b) "relationships with other groups" (ibid.) When people answer "yes" to the second question and "no" to the first one, it is a prerequisite (or, maybe, even *the* prerequisite) for a language shift. Even mixed families may fail to ensure a language shift, as it is in the case of Bará and Tuyuka discussed above.

There are different types of language shifts depending on how numerous the speakers of the winning language are, how long it takes to completely switch from the original language to the winning language, and to what extent the standard variant of the latter is available. The winning language, in this case, must be labeled "U" because it survives and undergoes contact changes, while the original language is "I" since it is the source of the influence.

Let us first consider a language shift to a majority language. Here, language U, spoken by a large group of people, is subjected to the influence of language I, which is either the language of the power elite, or the language of a small aboriginal population. In both cases, the people U are obliged to know language I (at least to a small extent), but are able to use language U in most situations. The people I value the knowledge of language U, and its standard variant is available for them. So, if the contact lasts long enough, they are able to learn most of the grammar of language U, but retain several lexemes that have no equivalents in language U (either the names of local flora, fauna, landscape details etc., or words that are necessary for dealing with the ruling establishment); native speakers of language U also tend to use these words (without inventing huge periphrastic constructions to name the corresponding objects). As a result, language U will acquire several borrowings in cultural vocabulary; if the contact situation lasts long enough, regular phonetic correspondences will be observed in these borrowings, but if the language shift is rapid, regularity will be violated.

In such a case, traces of grammatical interference are minimal, and basic vocabulary is not affected. Genes and myths will be mostly U with some admixture of I elements (although later, frequencies of the haplogroups may change as a result of genetic drift). Material culture will be mixed, since people U will have to adjust to either new local or new social conditions; most of it, however, will stem from U.

However, in the case of a shift to the minority language, the situation changes drastically.

Sometimes it happens that the linguistic minority is dominant, which leads to all the people of the region switching to its language (U). Naturally, the standard variant of U is not available for the majority of people I, so many of them acquire the language U from imperfect

speakers. Ö. Dahl speaks about "suboptimal" language transmission (Dahl 2004: 110, 200) in such situations.

In this case, it matters to what degree this lect is imperfect. Those who study second language acquisition mention that there is a typical sequence in which the elements of language are learned (see, e.g., Cook 2008): isolated words and formulae arrive first, then comes grammar, and, lastly, phonetics. As an example, V. Cook cites "Henry Kissinger, forever sounding German" (Cook 2008: 135). This is significant, because it predicts which mistakes in language U will be transmitted to children from parents who do not know language U well enough. Thus, "there are different degrees of suboptimality" (Dahl 2004: 110).

As for the expected sequence of learning different grammar elements, there are some works that study the sequence of learning English, showing that it does not depend much on the first language of the student (Cook 2008: 25–27 with the works she cites, esp. Hannan 2004), but much less is known about the sequence of learning grammar elements of other languages.

Empirical observations show that more transparent constructions are learned more eagerly than less transparent ones (cf. Ö. Dahl's opinion that mature language features "are highly prone to being filtered out in suboptimal language acquisition", Dahl 2004: 207). But what would be considered more transparent depends on the properties of the first language: for example, definiteness is transparent for speakers of Germanic languages, but absolutely obscure for Russian-speaking people; gender is quite transparent (and easily penetrates over the course of language contacts) for Indo-Aryans and Dravidians, but obscure for Finno-Ugrians. Regular models (irrespective of their being syntactic or derivational) are learned more effectively than irregular ones.

As a rule, the more an element of the second language looks like some similar element of the first language, the more quickly it is learned, thus having the better chance to pass from one language to another even if the contact is short. When languages are closely related, they have much in common, but even in distantly related (or totally unrelated) languages there may be similarities (including chance similarities), and they are learned more easily, e.g., the first English grammar morpheme learned by Spanish speakers is the plural marker *-s* (Cook 2008: 25), similar to the Spanish plural marker *-s*. Even when a pidgin is created, it may contain an element that is linked to both languages — lexifier and substrate: e.g., in Russenorsk there is a preposition *på* that is similar (in form as well as in meaning) both to Norwegian *på* and Russian *no* (cf. Broch & Jahr 1981; 1990, Perekhval'skaya 2008: 1999–201). In West African Pidgin English there is a quotative construction of the type *king i tok sey, yu don kom* 'the king said: "You have come"' (Hopper & Traugott 2008: 216); this construction may have been translated literally (i.e., "calqued") from Akan *se/si* 'that complementizer, quotative say'; or, possibly, represent a direct borrowing of Akan *se* (ibid.), but the word *sey* looks very much like English *say*, so that one might suggest the same situation as in the case of Russenorsk *på*.

When the degree of suboptimality is exceptionally high, contact is short and the standard variant of language U soon becomes unavailable, a "radical creole" (see Belikov 2006), or "abrupt creole" (see Thomason & Kaufman 1988) will arise. Since most of the elements of language U that people I got to learn were restricted to words and certain formulae, it shares the main part of the basic vocabulary with its lexifier (e.g., in Tokpisin, more than 80% of the Swadesh 100-word list have an English origin, see Belikov 2009: 65–66), but grammar and phonetics are totally different. Such formulae as "who is that" are considered to be single words: e.g., in Tokpisin the formula *who's that* became the word for 'who' (*husat*; cf. also Tokpisin *wanem* 'what' < Engl. *what name*).

Such contact situations are extremely rare, since the conditions that can produce them are very specific (see Burlak & Starostin 2005: 64–68): there must be a short contact period with a

highly prestigious language without mixed families and mixed settlements, so that the standard variant of that prestigious language is never fully available, and at the same time there is no possibility to switch to another language, the standard variant of which is available. Such a situation would be impossible in the Paleolithic period; historians say that the proper conditions for the formation of creole languages could not have arisen earlier than the arrival of the complex chiefdom stage (A. Davletshin, p. c.).

A creole language can be recognized by comparing its 100-word list with the lists of its lexifier and the other members of the lexifier's group (Burlak 2008). If several languages are normal descendants of a proto-language, the words in their 100-word lists are replaced randomly, so that a word present in only one language of its subgroup may have cognates among the descendants of the protolanguage outside this subgroup. E.g., Swedish words that have no Danish equivalents may find cognates in Germanic languages outside the North Germanic subgroup, and these cognates are distributed randomly, cf. *buk* 'belly' = Germ. *Bauch* (but not Engl. *belly*), *löf* 'leaf' = Engl. *leaf* (but not Germ. *Blatt*), *små* 'small' = Engl. *small* (but not Germ. *klein*), *stjert* 'tail' = Dutch *staart* (but not Engl. *tail* and Germ. *Schwanz*). However, creole words that have no equivalent in the lexifier language will never have cognates in the other languages of the lexifier's group, because the creole language differs from all of them (1) by the words that distinguish the lexifier language from the other languages in its group and (2) the words that distinguish the creole language from its lexifier.

When, already after the creolization (or without it), the standard variant of the U language becomes available, a post-creole (post-pidgin, respectively) continuum develops (see, e.g., O'Donnell & Todd 1980). Thus, the situation becomes similar to "dialect regrouping" (see below).

If the contact lasts longer, speakers of I have time to learn not only lots of words, but also some grammar of language U, although not completely.

If there are mixed families where one of the parents is a native speaker of language U, children, in their sensitive period, are introduced to (a) the standard variant of language U spoken by one of the parents, and (b) the substandard variant of language U spoken by the other one (sometimes, however, even both parents may speak the substandard variant of language U, if neither of them comes from the ethnic group U). Other children, whom they may play with, are mostly from such mixed families, and thus know the substandard variant of U. Hence, such children become native speakers of an imperfect language.

The phonetics of this lect will differ very much from the one of standard U, because (a) pronunciation is the most difficult part of the second language acquisition and (b) the phonetics of the first language is affected by competence in the second one (see above); my own experience testifies that even a week of intensive communication with people that exhibit a substandard form of pronunciation may suffice for words to begin to be stored in memory in substandard sound forms. When the sound system of language I differs considerably from the sound system of U, phonetic correspondences between Proto-U and the resulting lect will be regular but odd, cf. Arm. *erku* < PIE *$d\mu\bar{o}$ 'two' (for information on prehistoric contacts of Armenian, see Diakonov 1985), Hung. *valy* < PU *oδe 'year', Hung. *bogyo* < PU *pola 'berry', Toch. A *es* < PIE *omso- 'shoulder'; in Japanese, the syllable structure of Altaic was totally rearranged. In order to provide absolutely irregular phonetic correspondences, the population must be extremely small (as it was on the Norfolk island where the Norfuk creole language arose, see Belikov 1998: 56 for details).

In the case of language shift to the minority language, some phonemes or phonological oppositions may be lost, as were, for example, the oppositions between voiced and unvoiced, aspirated and non-aspirated consonants in Tocharian; in the opposite case, when language U

is the majority language, it tends rather to acquire additional phonemes and oppositions, as did Sanskrit, by adding the cerebral series of consonants (originally from Dravidian) to the Indo-European heritage.

The language that results from the shift to the minority language will not necessarily be simplified, because, in mixed families, the Primary Linguistic Data (primary linguistic input) will include phrases spoken by the native U-speaker, and this is a means to acquire a considerable part of the U grammar. When a child acquires a language he/she does not learn all its elements by heart, but rather constructs the grammar of his/her native language (see, e.g., Slobin 1971). This may seem paradoxical, but fragments of phrases produced by non-native speakers can make children reconstruct (or even construct) the whole rule. E.g., if a child knows that in the language that he/she is acquiring there are no agreement markers and the word order is NAdj and GenN ('sheep' + 'black' means 'black sheep' and 'friend's' + 'sheep' means 'friend's sheep'), he/she can easily deduce that 'friend's' + 'sheep' + 'black' must mean 'friend's black sheep', but even if both attributes appear in the same position (e.g., 'black' + 'sheep', 'friend's' + 'sheep'), the entire rule is also deduced unequivocally: 'friend's' + 'black' + 'sheep' (because 'black' + 'friend's' + 'sheep' would be ambiguous). This could, perhaps, be the reason that underlies the structure of creole languages, where the first generation of native speakers has little opportunity to hear long phrases that are governed by systemic rules.

Thus, the minority language that has been influenced as a result of shifting to it, will not necessarily be simplified, since children acquiring language are able not only to maintain complex features but also produce them (Dahl 2004; Trudgill 2011). It does not mean that all these complex features must be inherited from language U and may not be constructed *de novo*. E.g., when a non-Indo-European people learned an Indo-European dialect (which would later become a Proto-Tocharian), they managed to acquire a substantial part of verbal morphology (although the 2 and 3 Sg present tense endings *-t* and *-ṣ* respectively are hardly traceable directly back to PIE *-si* and *-ti*), but nominal declension was completely rebuilt. There are highly irregular, even suppletive verb paradigms in Tocharian, and some parts of these suppletive paradigms are not of Indo-European origin, e.g., Toch. A *kälk* 'he/she went', *naṣ* 'he/she is' (cf. *yäṣ* 'he/she goes' < PIE *i-*, *ṣeṣ* 'he/she was' < PIE *s-*). Maybe a similar situation took place in the Hittite Empire when Luvians did not master Hittite completely, and this may have been the reason of clitic reduplication in New Hittite (Yakubovich 2009).

In such a language, there will be a few basic words retained from language I (even belonging to the most stable half of the 100-word list); that, in fact, will be the words that (maybe by mere chance) managed to be more frequent than their synonyms of U origin. Pronouns are not infrequently retained in just such a way: e.g., when Polynesians of the Niuafo'ou island switched from their indigenous language (belonging to the Nuclear Polynesian group) to the Tongan language (belonging to the Tongic group, another branch of the Polynesian languages; see Collocott 1922, Dye 1980, Belikov 1989: 49), they retained their personal pronouns 'we [excl.]', 'you [du.]', 'you [pl.]' (and some other basic lexemes); maybe this was also the case with the Tocharian first person pronouns — Toch. A *näṣ* (masc.), *ñuk* (fem.), Toch. B *ñiś* (masc., fem.) — which can hardly be traced back directly to Proto-Indo-European *eĝhom* or *m-* (neither does the Tocharian A gender opposition). Also, some basic meanings will not be expressed by inherited roots, but rather derived (e.g., Toch. A *klots*, B *klautso* 'ear' does not go back to PIE *aus-* but is derived from PIE *k̂leus-* 'to hear', thus meaning something like 'hearing device'; Toch. A 'bird' is *salat lu*, literally 'jumping animal'). A similar case might be between Sumerian and Hurro-Urartian, although it is not known exactly whether a Hurro-Urartian tribe switched to Sumerian, or vice versa, a para-Sumerian ethnic group switched to Hurro-Urartian (Kassian, forthc.)

Summing up, in the case of a shift to the minority language the resulting language U acquires a lot of loanwords, even in basic vocabulary; likewise, some basic words may be replaced by derivatives, phonetic correspondences are odd, grammar interference is abundant. Genes, however, are expected to be mostly I, cf. e.g., the situation with Hungarian and Hungarians: the Hungarian language belongs to the Finno-Ugric family while the Hungarian people share the most part of their genetic material with neighboring Slavic peoples. "It is generally accepted that Magyars (Hungarians) were strong enough to mix with preceding population of this region and make them speaking Hungarian language, but they were not *numerous* enough to make recordable contribution into their gene pool. This pattern of invasion with language change but without change in gene pool is called *elite dominance model*" (Balanovsky et al. 2013). Material culture will be mixed (U traits will prevail).

It should be noted that, although the results of both types of language shift are often called "substrate", these "substrates" may actually have different natures.

An intermediate case is the Greek language: there are a few loanwords in its 100-word list (σελήνη 'moon', ξανθός 'yellow') as well as some derivatives (ὄφθαλμος 'eye'), some odd phonetic correspondences (PIE *s-, *i̯-, *u̯- > Gr. h-, PIE *-m > Gr. -n), many loanwords in the cultural sphere (θάλασσα 'sea' and many others), but the Indo-European grammar is preserved.

The last type of contacts may be called dialect regrouping. It may also occur between closely related (mutually intelligible, "obviously related", see Burlak & Starostin 2005: 19) languages. When speakers of both dialects are numerous, there are mixed settlements and mixed families, and nobody cares to keep his/her language pure and different from the other one, a *koiné* is composed (which may later split into another group of dialects). This koiné includes both vocabulary (even basic vocabulary) and grammar forms from both sources. Myths, genes and material culture will be common. An example of such contact is the history of North Slavic (the language of Medieval Novgorod) and Standard Old Russian. The northern branch separated from the Common Slavic in times preceding the second palatalization, it had its own NSg marker *e* in the *ŏ-declension instead of Common Slavic *ъ and several other peculiarities. But in the Middle Ages it began to converge with Standard Old Russian, and the results of this can be seen not only in Novgorod but in Moscow as well: in Modern Russian, there are multiple grammatical forms of North Slavic origin (see Zalizniak 2012; forms ancestral to Modern Russian are shadowed):

	Novgorod	Standard Old Russian
N Sg	*городе*	*городъ*
Gen Sg	*у сестрѣ*	*у сестры*
Loc Sg	*в землѣ, на конѣ*	*в земли, на кони*
Nom Pl	*моѣ коровѣ*	*мои коровы*
Nom Du	*два села*	*двѣ селѣ*
Imp Sg	*помоги*	*помози*
Imp Pl	*везите*	*везете*
3 Sg Pres	*везе*	*везетъ*
Ger	*везя*	*веза*

Such a situation, when grammar forms are acquired from both sources in equal measure, cannot be described either as "borrowing" or as "language shift"; this is a real convergence.

This type of evolution, obviously, is quite common not only in linguistics but also in biology, where subspecies and closely related species can diverge, then hybridize, then diverge and hybridize again and so on (Iordanskiy 2001: ill. 10). But it must be specified that both in linguistics and in biology only closely related forms can evolve in such a way. In fact, there is nothing surprising about it: individuals of the same species, as well as individual variants of the same language can mix, and this ability is lost only when languages have become mutually unintelligible and species have acquired some mechanisms of pre-zygotic or post-zygotic isolation, and both processes require time.

Naturally, it should be noted that the described types are not rigid: they may evolve, so that one type may change into another, and one population may pass through several contacts of different types. Thus, Sanskrit shows that during the original contact period it was a majority language: no borrowings in the 100-word list, multiple loanwords in cultural vocabulary, several loan phonemes, PIE phonetic inventory and grammar mainly preserved; this agrees with the hypothesis attributing early Indo-Aryan presence to the Gandhara grave culture (see Mallory 1989): around 1900 BC, many Punjab settlements became deserted, so that the people of the Gandhara grave culture, appearing one century later in the Swat valley, were likely to form a linguistic majority. On the contrary, modern Indo-Aryan languages show traces of being minority languages: cf., e.g., the odd phonetic correspondences between Lahnda *sijjh* and Sanskrit *súvar*, *surya-* 'sun', Hindi *ā̃kh* and Sanskrit *ákṣmi-* 'eye'; both Hindi and Lahnda have several loanwords in their 100-word lists, and their grammar differs much from the Sanskrit (resp. PIE) one. At the same time, for instance, Russian on its way from Proto-Slavic to modern times had undergone various contacts, but was never a minority language, and it retains more grammar elements from Proto-Slavic (and the phonetic correspondences are trivial as well).

I believe that it is possible to define particular language elements that can shed light on the prehistoric demographical situation by means of comparative linguistics. And this, in turn, would help to reconcile linguistic, archaeological and genetic data and produce a more detailed picture of the history of the peoples and their languages, and, maybe, make it possible to localize the Proto-Indo-European homeland.

Literature

Balanovsky 2008 — O. P. BALANOVSKY. Genofond vysokikh shirot Evrazii ili otkuda prishli saamy? In: A. A. Velichko & S.A. Vasil'ev (eds.). *Put' na sever: okruzhayuschaya sreda i samye rannie obitateli Arktiki i Subarktiki (materialy mezhdunarodnoy konferencii)*. Moscow, 2008. P. 277–282.

Balanovsky et al. 2013 — O. BALANOVSKY, O. UTEVSKA, E. BALANOVSKA. Genetics of Indo-European populations: the past, the future. In: *this volume*. P. 23–36.

Belikov 1989 — V. I. BELIKOV. Drevnejshaya istoriya i real'nost' lingvogeneticheskikh dendrogramm. In: *Lingvisticheskaya rekonstrukciya i drevneyshaya istoriya vostoka*: *materialy k diskussiyam na Mezhdunarodnoy konferencii (Moskva, 29 maya — 2 iyunya 1989 g.)*. Vol. 1. Moscow, 1989. P. 44–54.

Belikov 1998 — V. I. BELIKOV. *Pidzhiny i kreol'skie yazyki Okeanii*. Moscow, 1998.

Belikov 2006 — V. I. BELIKOV. *Konvergentnye processy v lingvogeneze*. Thesis Doct. phil. nauk. Moscow: Lomonosov Moscow State University, 2006.

Belikov 2009 — V. I. BELIKOV. Language Contacts and Genealogical Classification of Languages. In: *Journal of Language Relationship*, No. 1 (2009). P. 49–68.

Beryozkin 2008 — Yu. E. BERYOZKIN. Sibirsko-saamskie svyazi v oblasti mifologii na fone syuzheta ATU 480. In: *Natales grate numeras?* St. Petersburg, 2008. P. 119–143.

Beryozkin n.d. — Yu. E. BERYOZKIN. *Tematicheskaya klassifikaciya i raspredelenie fol'klorno-mifologicheskikh motivov po arealam: Analiticheskiy katalog*. Available at: http://www.ruthenia.ru/folklore/berezkin/ [accessed: 10.11.2012].

Blench & Spriggs 1997 — R. M. BLENCH & M. SPRIGGS (eds.). *Archaeology and Language*. 1: *Theoretical and Methodological Orientations.* London, 1997.

Blench & Spriggs 1998 — R. M. BLENCH & M. SPRIGGS (eds.). *Archaeology and Language*. 2: *Correlating Archaeological and Linguistic Hypotheses.* London, 1998.

Blench & Spriggs 1999a — R. M. BLENCH & M. SPRIGGS (eds.). *Archaeology and Language*. 3: *Artefacts, Languages, and Texts.* London, 1999.

Blench & Spriggs 1999b — R. M. BLENCH & M. SPRIGGS (eds.). *Archaeology and Language*. 4: *Language Change and Cultural Transformation.* London, 1999.

Broch & Jahr 1981 — I. BROCH, E. H. JAHR. Russenorsk — et pidginspråk i Norge. In: *Tromsø-studier i språkvitenskap*, vol. 3. Oslo, 1981.

Broch & Jahr 1990 — I. BROCH, E. H. JAHR. Russenorsk: the Russo-Norwegian pidgin. New findings. In: *Tromsø studies in linguistics*, vol. 11. Oslo, 1990. P. 62–68.

Burlak 2008 — S. A. BURLAK. Kreol'skie yazyki I glottokhronologiya. In: *Orientalia et classica*, vol. 19. Moscow, 2008. P. 499–508.

Burlak & Starostin 2005 — S. A. BURLAK, S. A. STAROSTIN. *Sravnitel'no-istoricheskoe yazykoznanie* [*Comparative Linguistics*]. Moscow, 2005.

Collocott 1922 — E. E. V. COLLOCOTT. The speech of Niua Fo'ou. In: *The Journal of the Polynesian Society*, Vol. 31, No. 4(124) (1922). P. 185–189.

Cook 2008 — V. COOK. *Second Language Learning and Language Teaching*. 4th ed. London, 2008.

Cook & al. 2007 — V. COOK, K. AL-EBARY, B. AL-GARAWI, C.-Y. CHANG, K. HUANG, S. LEE, G. LI, H. MITANI, Y.-C. SIEH. *Effects of cue processing in the Competition Model on the first language of L2 users*. Paper presented at the 17th EUROSLA annual conference, Newcastle upon Tyne. 2007.

Dahl 2004 — Ö. DAHL. *The Growth and Maintenance of Linguistic Complexity*. Amsterdam; Philadelphia, 2004.

Diakonov 1985 — I. M. DIAKONOV. Hurro-Urartian Borrowings in Old Armenian. In: *Journal of the American Oriental Society*, Vol. 105, No. 4, 1985. P. 597–603.

Dye 1980 — T. S. DYE. The linguistic position of Niuafo'ou. In: *The Journal of the Polynesian Society*, Vol. 89, No. 3 (1980). P. 349–357.

Forster & Renfrew 2011 — P. FORSTER & C. RENFREW. Mother tongue and Y chromosomes. In: *Science*, vol. 333, 2011. P. 1390–1391.

Gardner 2007 — R. GARDNER. The socio-educational model of second language acquisition. In: *EUROSLA Yearbook*, vol. 6, 2007. P. 237–260.

Gumpertz & Wilson 1971 — J. J. GUMPERTZ, R. WILSON. Convergence and Creolization: A Case from the Indo-Aryan / Dravidian Border in India. In: D. Hymes, ed. *Pidginization and Creolization of Languages*. Cambridge, 1971. P. 151–167.

Gurvich 1968 — I. S. GURVICH. Kul't svyaschennykh kamney v tundrovoy zone Evrazii. In: *Problemy antropologii i istoricheskoy etnografii Azii*. Moscow, 1968. P. 230–239.

Hannan 2004 — M. HANNAN. *A study of the development of the English verbal morphemes in the grammar of 4–9 year old Bengali-speaking children in the London borough of Tower Hamlets*. PhD thesis, University of Essex, 2004.

Hartsuiker et al. 2004 — R. J. HARTSUIKER, M. J. PICKERING, E. VELTKAMP. Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. In: *Psychological Science*, Vol. 16, No. 6, 2004. P. 409–414.

Hopper & Traugott 2008 — P. J. HOPPER, E. C. TRAUGOTT. *Grammaticalization*. 2nd ed. Cambridge, 2008.

Iordanskiy 2001 — N. N. IORDANSKIY. *Evol'uciya zhizni*. Moscow, 2001.

Jackson 1974 — J. JACKSON. Language Identity of the Colombian Vaupes Indians. In: R. Bauman & J. Sherzer, eds. *Explorations in the Ethnography of Speaking*. Cambridge, 1974. P. 50–64.

Kassian, forthc. — A. KASSIAN. Lexical matches between Sumerian and Hurro-Urartian: possible historical scenarios. In: *Language Dynamics and Change*, forthcoming volume.

Kossinna 1911 — G. KOSSINNA. *Die Herkunft der Germanen. Zur Methode der Siedlungsarchäologie*. Leipzig, 1911.

Kuz'menko 2011 — Yu. K. KUZ'MENKO. *Rannie germancy i ikh sosedi: Lingvistika, arkheologiya, genetika* [*Early Germans and their neighbours: Linguistics, archaeology, genetics*]. St. Petersburg, 2011.

Khit' 1983 — G. L. KHIT'. *Dermatoglifika narodov SSSR*. Moscow, 1983.

Khit' 1991 — G. L. KHIT'. Saamy v dermatologicheskoy sistematike finno-ugrov Evrazii. In: *Proiskhozhdenie saamov* [*Origin of Saami*]. Moscow, 1991. P. 59–82.

Levontina 2010 — I. Levontina. *Russkiy so slovarem*. Moscow, 2010.

Locastro 1987 — V. Locastro. Aizuchi: a Japanese conversational routine. In: L. E. Smith, ed. *Discourse Across Cultures*. New York, 1987. P. 101–113.

Mallory 1989 — J. P. Mallory. *In Search of the Indo-Europeans: Language, Archaeology, and Myth*. London, 1989.

McWhorter 2005 — J. H. McWhorter. *Defining creole*. Oxford, 2005.

Metspalu & al. 2011 — M. Metspalu, I. Gallego Romero, B. Yunusbayev, G. Chaubey, Ch. B. Mallick, G. Hudjashov, M. Nelis, R. Mägi, E. Metspalu, M. Remm, R. Pitchappan, L. Singh, K. Thangaraj, R. Villems, T. Kivisild. Shared and Unique Components of Human Population Structure and Genome-Wide Signals of Positive Selection in South Asia. In: *The American Journal of Human Genetics*, vol. 89, no. 6, 2011. P. 731–744.

Nikolaeva 2011 — N. A. Nikolaeva. *Etno-kul'turnye processy na Severnom Kavkaze v III–II tys. do n.e. v kontekste drevney istorii Evropy I Blizhnego Vostoka*. Moscow, 2011.

Nichols 1992 — J. Nichols. *Linguistic diversity in space and time*. Chicago, 1992.

O'Donnell & Todd 1980 — W.R. O'Donnell & L. Todd. *Variety in contemporary English*. London, 1980.

Peiros 1997 — I. Peiros. Are correlations between archaeological and linguistic reconstructions possible? In: R. Blench, & M. Spriggs, eds. *Archaeology and Language, vol. I: Theoretical and Methodological Orientations*. London, 1997. P. 149–157.

Perekhval'skaya 2008 — E. V. Perekhval'skaya. *Russkie pidzhiny*. St. Petersburg, 2008.

Rankin 2003 — R. L. Rankin. The Comparative Method. In: B. D. Joseph, R. D. Janda, eds. *The Handbook of Historical Linguistics*. Oxford, 2003. P. 183–212.

Sahoo & al. 2006 — S. Sahoo, A. Singh, G. Himabindu, Jh. Banerjee, T. Sitalaximi, S. Gaikwad, R. Trivedi, P. Endicott, T. Kivisild, M. Metspalu, R. Villems, V. K. Kashyap. A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios. In: *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 4, 2006. P. 843–848.

Sengupta & al. 2006 — S. Sengupta, L. A. Zhivotovsky, R. King, S. Q. Mehdi, C.A. Edmonds, C. E. Chow, A. A. Lin, M. Mitra, S.K. Sil, A. Ramesh, M. V. Usha Rani, C. M. Thakur, L. L. Cavalli-Sforza, P. P. Majumder, P. A. Underhill. Polarity and Temporality of High-Resolution Y-Chromosome Distributions in India Identify Both Indigenous and Exogenous Expansions and Reveal Minor Genetic Influence of Central Asian Pastoralists. In: *American Journal of Human Genetics*, vol. 78, no. 2, 2006. P. 202–221.

Sankoff 2001 — G. Sankoff. Linguistic Outcomes of Language Contact. In: P. Trudgill, J. Chambers & N. Schilling-Estes, eds. *Handbook of Sociolinguistics*. Oxford, 2001. P. 638–668.

Sharma & al. 2009 — S. Sharma, E. Rai, P. Sharma, M. Jena, Sh. Singh, K. Darvishi, A. K. Bhat, A. J. S. Bhanwer, P. K. Tiwari, R. N. K. Bamezai. The Indian origin of paternal haplogroup R1a1* substantiates the autochthonous origin of Brahmins and the caste system. In: *Journal of Human Genetics*, vol. 54, 2009. P. 47–55.

Simchenko 1976 — Yu. B. Simchenko. *Kul'tura okhotnikov na oleney Severnoy Evrazii*. Moscow, 1976.

Slobin 1971 — D. I. Slobin. *Psycholinguistics*. Glenview, 1971.

Thomason 2008 — S. Thomason. *Does language contact simplify grammars?* Paper presented at the Deutsche Gesellschaft für Sprachwissenschaft, Bamberg, February 2008. Available at:
http://www-personal.umich.edu/~thomason/temp/simple2.pdf [accessed: 10.11.2012].

Thomason & Kaufman 1988 — S. G. Thomason, T. Kaufman. *Language contact, creolization and genetic linguistics*. Berkeley; Los Angeles; Oxford, 1988.

Trudgill 2004 — P. Trudgill. Linguistic and social typology: the Austronesian migrations and phoneme inventories. In: *Linguistic Typology*, vol. 8, 2004. P. 305–320.

Trudgill 2011 — P. Trudgill. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford, 2011.

Underhill & al. 2009 — P. A. Underhill, N. M. Myres, S. Rootsi, M. Metspalu, L. A. Zhivotovsky, R. J. King, A. A. Lin, C.-E. T. Chow, O. Semino, V. Battaglia, I. Kutuev, M. Järve, G. Chaubey, Q. Ayub, A. Mohyuddin, S. Q. Mehdi, S. Sengupta, E. I. Rogaev, E. K. Khusnutdinova, A. Pshenichnov, O. Balanovsky, E. Balanovska, N. Jeran, D. H. Augustin, M. Baldovic, R. J. Herrera, K. Thangaraj, V. Singh, L. Singh, P. Majumder, P. Rudan, D. Primorac, R. Villems, T. Kivisild. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. In: *European Journal of Human Genetics*, vol. 18, 2009. P. 479–484.

Yakubovich 2009 — I. Yakubovich. *Sociolinguistics of the Luvian Language*. Leiden; Boston, 2009.

Zalizniak 2012 — A. A. Zalizniak. *Ob istorii russkogo yazyka* (Lecture, February, 24) Moscow, 2012. Available at:
http://www.mumi-troll.ru/secondary/lections/0029.html [accessed: 10.11.2012].

С. А. Бурлак. Языки, геном, родство и контакты.

В статье производится классификация языковых контактов на основании того, какие различия можно усмотреть в их результатах средствами сравнительно-исторического языкознания. Выделяется несколько типов билингвальных контактов и языковых сдвигов; они различаются по влиянию, оказываемому ими на словарь, грамматику и фонетику подвергшихся им языков. Таким образом, видя в современном языке следы того или иного типа контактов, можно сделать вывод о той этнолингвистической ситуации, в которой этот контакт происходил. Такого рода информация о доисторических обществах может помочь увязать лингвистические данные с археологическими и генетическими, для того чтобы получить более детальную картину истории народов и их языков.

Ключевые слова: языковое родство, языковые контакты, языковой сдвиг.