PROCEEDINGS

of

the 12th Seoul International Altaistic Conference

Multidimensional understandings
of languages and cultures of Altaic people

Dates: 16~19 July 2015

Venue: Shinyang Hall II, Seoul National Univ., Korea

The Altaic Society of Korea
Institute of Altaic Studies, SNU

Reconstruction of Old Kipchak

Henryk JANKOWSKI
Adam Mickiewicz University in Poznań, Poland

1. Old Kipchak, Middle Kipchak and modern Kipchak languages

The goal of this paper is to consider some principles of the establishment of those Old Kipchak words which are distinctive from other Old Turkic languages. However, the reconstruction of a more extensive lexicon should also include basic words most of which are identical with corresponding words of other Old Turkic languages.

Old Kipchak is a language that was spoken in two periods, one Early Old Kipchak from the late-8th century when it first appeared (Golden 2014: 183–184) to the mid-11th century, the other Late Old Kipchak from the second half of the 11th century to the Mongol invasion in the 1230s. There is only scanty documentation of Old Kipchak from its late period by Maḥmūd al-Kašġārī, often ambiguous, the remaining being single glosses in Old Russian, Arabic and other documents. The next period in the development of this language is Middle Kipchak, dated to the 13rd–16th centuries (Berta 1998: 158). It is customary to speak of the emergence of individual modern Turkic languages from the 16th century onwards. This periodisation is not without problems. Firstly, the Bashkirs who speak a Kipchak

language are known already in 922 in Ibn Faḍlān's report (Kmietowicz, Kmietowicz, Lewicki 1985: 87, 97), but it is unknown if their first Turkic language was of the Kipchak type and only lately Kipchakised. The non-Kipchak layer in Bashkir lexicon may be an adstrate, but also a substrate. A similar case are the Karakalpaks, who – as such – are known from the 16th century, but they may be connected with the *Černye Klobuki* 'Black Cowls', mentioned in Old Russian chronicles in the 12th century and elsewhere (Golden 1992: 403). In fact, the non-Kipchak features in their languages, such as *d-*, e.g. *diz-* 'to string, to set', *düz* 'plain; steppe', and *g-*, e.g. *gez-* 'to wander, to roam', can be explained by both later Turkmen influence and an earlier substrate. Lastly, Armeno-Kipchak documented in the 16th–17th centuries (Pritsak 1959: 74) reveals many old lexical features and cannot be called a modern language. A similar case is Karaim, still a living language. Although documented at a relatively late period, it shows many archaic features typical of Middle Kipchak. Recently, two Karaim poems written in the 16th-century have been discovered in a 17th-century manuscript (Jankowski 2014).

Two basic groups of sources for Middle Kipchak are Mamluk-Kipchak documents[1] and *Codex Cumanicus*.[2] Material of Kumans who settled in Hungary in the 13th century, although fragmentary and recorded at a much later time, is also an important source.[3] A third group is West Karaim and old Crimean Karaim and a fourth one is Armeno-Kipchak (*xïpčax tili, tatarča*). Unfortunately, as İnan (1953: 56–64) has proved, the Arab compilers of compendia for Mamluk-Turkic of Mamluk-Kipchak did not have a clear idea of the differences between Oghuz (which they termed Turkmen) and Kipchak, see also Zajączkowski's (1954: 21) remarks on Oghuz and Kipchak words in Bul., e.g. *išïr-* and *tišlä-* 'to bite', *yaz-* and *čïz-* 'to write', and phonetic variants typical of Oghuz and Kipchak, e.g. *duz* and *tuz* 'salt' (Zajączkowski 1938: 20, 53). Therefore, almost all documents of this kind evidence words and forms of both Oghuz and Kipchak character. Clauson (1972: xxiv) was right when he said that the only purely Kipchak document among medieval Kipchak sources in Codex Cumanicus. The same must be said of Hungarian Kipchak, but – as said above – the available material is scarce and late.

[1] There is much literature on Middle Kipchak (Zajączkowski 1938: v–vi, Atalay 1945: x–xi, İnan 1953, Kuryšžanov 1970: 13–26, Fazylov, Zijaeva 1978: 5–10, Eckmann 1982–1983, Öztopçu 1989: 2–4, Toparlı, Vural, Karaatlı 2007: v–ix). Most Arabic dictionaries, grammars and practical aims for learning Kipchak, Turkic Kipchak or Turkic were composed or copied in Egypt and Syria, and only Ibn Muhanna's dictionary – the language of which is debatable – and Codex Cumanicus were written elsewhere.

[2] From the very abundant literature on Codex Cumanicus we can show some most significant studies such as Kuun (1880), Grønbech (1942), Gabain (1959) and Drimba (2000).

2. Principles of reconstruction

Old Kipchak has been the object of study in only a few papers. Some authors examined Kipchak (*Polovets*) words in Old Russian documents (Zajączkowski 1949, Menges 1951) and a paper was devoted to the Kipchak words in Maḥmūd al-Kašġārī's dictionary (Alimov 2014).

Zajączkowski examined the following Kipchak male names, which we can regard as distinctive from Oghuz: *Boluš* (1054) 'help' (pp. 34–35), *Qurtqa* (1103) 'old woman' (pp. 37), *Itler* (1095) 'dogs' (pp. 37, 41), *Turuntay* (1185) 'sparrow-hawk' (pp. 42–43), *Targołowe* < *taryïl* 'speckled', *Sevenč* (1151) 'like; joy' (pp. 47) (not *Söyünč*!).

[3] The basic study is Mándoky Kongur (1993) in which *ca* 110 words are collected. Róna-Tas and Berta (2011: 1340–1341) list thirty-five Kuman words borrowed into Hungarian, basing largely on Mándoky Kongur.

Alimov (2014: 149–152) has demonstrated fifty-seven words which Kašġārī qualified as Kipchak. Unfortunately, many of them are not distinctive, some being provided by the author as used by both the Oghuz and the Kipchaks, and sometimes by still other Turkic peoples. The words indicated by Kašġārī as only Kipchak, are the following: *özle* 'noon', *apa* 'bear', *učan* 'ship with two sails', *ükil* 'much', *ketü* 'paralysed (hand)', *qoru* 'prickly hedge', *quzyïr-* 'come pouring and swirling (snow)', *töz-* 'suffer cold and hunger' (DLT 282–283). Of these words, *učan* is also attested in Old Russian (Zajączkowski 1949: 25), *ükil*, and *töz-* has survived in NW Turkic (ED 572).

Many important observations on the historical phonology of Kuman have been made in Grønbech (1942: 6–21).

There was an attempt at the reconstruction of some components of "Proto-Kipchak" by Musaev and Čečenov (2002: 216–338), but their considerations require revision, e.g. the periodisation with "Early Proto-Kipchak (the 10th BCE–2nd CE centuries)", "Middle Kipchak (the 2nd–7th centuries)"[4] and "Late Proto-Kypchaq (the 7th–12th centuries)" which is arbitrary and lacks reasonable grounds. It is unknown how Musaev and Čečenov would like to show the transition from Proto-Kipchak to modern Kipchak languages when they assume that the process of the split of Proto-Kipchak into four groups had started at the turn of the Christian era (2002: 217). Their treatment of the Pecheneks as of a Kipchak people is also ungrounded. However, the lexical comparisons by those authors are quite useful.

Recently Özyetgin (2001) has examined the verbs in Abū Ḥayyān's dictionary. She provided Old Turkic, contemporary and modern

[4] *srednekypčakskij*; an evident mistake for *sredneprakypčakskij* 'Middle Proto-Kipchak'.

equivalents of 724 analysed verbs and sought a Proto-Turkic reconstruc-reconstruction of each verb.[5]

In my opinion it is better to speak of Old Kipchak, since this language clearly existed, but is only fragmentarily evidenced and should be reconstructed. Therefore, we may speak of Proto-Kipchak as spoken before the 8th century, but its reconstruction is impossible at this stage of knowledge, and probably not important to do. Anyway, the attempt to reconstruct Old Kipchak must precede the reconstruction of Proto-Kipchak. The establishment of Old Kipchak roughly corresponds to the time of Old Turkic when it was first evidenced. Recently, some scholars distinguish East Old Turkic evidenced in runic inscriptions and Old Uyghur from West Old Turkic, and Róna-Tas and Berta (2011) reconstructed part of its lexis on the basis of Bulgar loanwords in Hungarian, e.g. WOT **uk* 'mind, understanding', cf. Hun. *ok* 'cause, reason', as an non-evidenced nominal equivalent of the verb *uq-*; WOT **ökür* 'ox', cf. Hun. *ökör* 'id' (Róna-Tas, Berta 2011: 629–631, 663–667). It is probably a single endeavour to reconstruct an Old Turkic language, since the reconstruction is not necessary for the attested ones, e.g. East Old Turkic and Karakhanid. Such a reconstruction should also be done for Old Oghuz, which is not identical to Old Turkic. For the others such as Pechenek, Khazar and Old Kirghiz, there is no sufficient material.

As for Proto-Turkic, it must be stressed that this language, not reconstructed yet in a systematic, comprehensive way, is dated to the third (Ramstedt 1957: 15) or forth millennium BC (Róna-Tas 1998: 68). However, Róna-Tas often prefers the use of the term Ancient Turkic rather than Proto-Turkic (1991: 30), even if he implicitly postulates the need to

[5] There are also many articles on various aspects of Kipchak, which cannot be quoted in this paper. Some important critical editions of Kipchak documents are shown in the abbreviations (Kipchak documents). There is also a grammar of MK (Karamanlıoğlu (1994), written in 1958, and a dictionary (Toparlı, Vural, Karaatlı 2007). This dictionary is very helpful, but each entry must be checked in the original source.

reconstruct Proto-Kipchak (1998: 67), without saying anything further about this hypothetic language.

Old Kipchak, as we can conclude from Middle Kipchak evidence, was probably very similar to other Old Turkic languages, including East Old Turkic, except Old Bulgar. Kašġārī's inclusion of the Kipchaks among those Turks who "speak pure Turkic" (DLT 83, facsimile 25) should be interpreted as the regarding of their language as not much different from the others. Their close relations to the Oghuz (İnan 1953: 53)[6] should also strengthen language contact and the similarity of both languages. Therefore, we have to reckon with various non-Kipchak influences in Old Kipchak similar to later Kipchak influence on other Turkic languages such as southern Altay and especially Kirghiz. On the other hand, to begin with the 15th century, Middle Kipchak in the Mamluk state gradually lost its importance in favour of Oghuz (Eckmann 1964, Öztopçu 2002: 10).

The reconstruction must be lexical, since the phonological criteria are not quite clear in written sources. Kipchak vocabulary should be contrasted with non-Kipchak, principally Oghuz, as most similar to Kipchak. However, as said above, Middle Kipchak was also very close to the other common Middle Turkic languages, most importantly, to Khwarezmian Turkic. Eckmann (1964: 35–36) shows only six distinctive features differentiating Kipchak from Kharezmian.

The following phonological questions should be raised: (1) if high vowels /i ï u ü/ had already a tendency to be reduced; (2) if the shift /e- o- ö-/ > /ie- wo- wö-/ already started; (3) if some cases of the change /y-/ > /ǧ-/ occurred in this period, as Kašġārī claimed. It is likely that other

[6] The same holds true of the Kuman-Uz relations (Marquart 1914: 30). In this article, I do not distinguish the Kipchaks from the Rus. *polovtsy*, Western *Qumans* nor the Eastern *Qūn*. Although the latter are reported to had originally inhabited China on the easternmost edge of the Turkic world (Marquart 1914: 39, 54), in the course of the 13th century they united with the Kimeks and Kipchaks and became one people (Mándoky Kongur 1993: 11).

changes which later made Kipchak more different from Oghuz and Khwarezmian, e.g. *-Vɣ* > *-Vv* in monosyllabic words, *-VɣV-* > *-VvV-*, *-Iɣ* > *-Iv* > *-Uv* and *b-* > *m-* before *n, ŋ* at the end of the syllable or in further syllables started occurring only at the end of the Late Middle Kipchak period. The change *y-* > *ǧ-*, signalled by Kašġārī, though in an ambiguous way (DLT 84, facsimile 26) is perhaps most enigmatic. In fact, Kašġārī's most examples are *y-* forms and *ǧ-* does not appear in Middle Kipchak at all. An especially difficult problem is whether OT *-d-, -d* has really shifted to *-z,-z* or *-ð-, -ð*, as Kašġārī reported. If it had changed, the further change into *-y-, -y* could not have been possible, since the change *z* > *y* is hardly possible. However, since some documents evidence the change to *-z,-z* or *-ð-, -ð*, it is reasonable to assume this change only in some minor variants of Kipchak, which later adopted *-y-, -y* of the major variant. At the same time one has to reckon with the influence of the literary tradition Karakhanid in which this shift is well known and established.

Some problems have already been solved by Grønbech. On the basis of such double notations in Codex Cumanicus as *ešit-* ~ *išit-* 'to hear' and *ber-* ~ *bir-* 'to give', he came to the conclusion (Grønbech 1942: 7, 12) that the OT close *e* was in Kuman retained, although he did not distinguish it from the open *ä*. Grønbech (1942: 20) also claimed that vowel length was preserved, though it was not reflected in his transcription, either. At this point one has to note that some authors (e.g. Zajączkowski, Öztopçu) consistently distinguished *e* from *ä*, whereas others did not (e.g. Atalay, Özyetgin). However, while Grønbech's rendering of the close *e* with *e* is correct and acceptable, the rendering of close *e* with the letter *i*, as in Toparlı's studies, is wrong. As for vowel length, there are only scarce indications to this, e.g. *beyš* 'five' in Qav. (73b, Toparlı, Çögenli, Yanık 1999). However, there are no sufficient grounds in our opinion to postulate the existence of vowel length in OK and, therefore, we do not use this feature.

Grønbech (1942: 9) also observed such variants as *tay* ~ *tav* 'mountain' and *teg-* ~ *tey-* 'to reach, to attain', *ipek* ~ *yibek* 'silk', *sučul-* ~ *čučul-* 'to undress'.

3. Reconstruction

The reconstruction of lexicon, with a few exceptions like personal pronouns, should be directed to non-basic lexicon, since basic vocabulary is in most cases identical in common Old Turkic languages, except for Old Bulgar or Old West Turkic, e.g. OT *tīz* 'knee', but WOT **tīr*. It is especially evident when we consider monosyllabic verbs which are identical, e.g. *al-* 'to take', *öl-* 'to die', *qal-* 'to stay', *ye-* 'to eat', or which differ only phonetically, e.g. OK *bar-* 'to go' and Old Turkish *var-* ~ *bar-*; OK *kėt-* 'to go (away)' and Old Turkish *git-*.

It seems reasonable to reconstruct such words of the basic lexicon as personal pronouns, e.g. *men* ~ *mėn* 'I' and reflexive pronouns, e.g. *kensi* '-self'.

Naturally, when sifting the words to reconstruct, one has to consider if a given word can be distinctive from its equivalents in other languages. From Brands's (1973: 16) true Turkic words, the following are diagnostic, *bol-* 'to be, to become', *ïγač* 'tree', *küz* 'autumn', *su(b/v)* 'water', *tay* 'mountain' and *yïldïz* 'star'.

It should be stressed that some words may be unknown in SW Turkic because of insufficient documentation. Some words which seem to be absent from Turkish and the other SW languages could once have been used, as we see from historical and dialect dictionaries and some petrified sayings, e.g. *bet* 'face', for the SW rests see ED 296, ÈSTJa II 121–122, or *ėrin* 'lip' (ED 223–233, ÈSTJa I 292–293).

3.1. Basic lexicon

There are a few lists of basic vocabulary, Swadesh's lists of 100 words (Swadesh 1972: 283), Yakhontov's list of 35 words based on Swadesh's 207, and Holman et alii's list (2008: 351) based on Swadesh's 100 list with 40 asterisked highly frequent words. Yakhontov 35 and Holman 40 lists are similar. As said above, most basic words are distinguished phonetically, only some items differ lexically.

The following words on Holman's list are distinctive from Old Oghuz and, therefore, diagnostic for Late Old Kipchak:

'two' OO *iki : OK *eki > eki ~ i͡eki
'I' OO *ben : OK *men
'water' OO *su : OK *suv
'liver' OO *baγïr : OK *bawïr > bawïr, bawur
'eye' OO *göz : OK *köz
'hand' OO *el : OK *qol
'hear' OO *išit- : OK *ešit-
'tree' OO *aγač : OK *aγač > aγač, aγaš, but also *terek, the basic word in some modern SW languages, e.g. CTat.
'fish' OO *bālïq : OK *balïq, but 'small fish' OO *čapaq : OK *čabaq
'name' OO *āt : OK *at
'stone' OO *dāš : OK *taš
'breasts' OO *gögüs : OK *kögüz, but also kökrek and OO *dõš : OK *töš;
'tooth' OO *dïš : OK *tiš
'bone' OO *süŋek; *kemik : OK *süŋük, süvük, süŋek, süyek, süvek
'tongue' OO *dil : OK *tïl
'skin' OO *deri : OK *teri
'night' OO *gēğe : OK *keče; OO *dũn : OK *tün
'horn' OO *boynuz : OK *müyüz, müŋüz
'knee' OO *dīz : OK *tiz > tiz, tize
'nose' OO *burun : OK *murun
'full' OO *dōlu : OK *tolu

'come' OO *gel- : OK *kel-
'star' OO *yïldïz : OK *yulduz > yulduz, ğulduz
'mountain' OO *dāɣ : OK *taw
'fire' OO *ōt : OK *ot
'see' OO *gör- : OK *kör-
'new' OO *yeŋi > yeŋi, yeni , OK *yaŋï, yaŋa > yaŋa, ğaŋa
'dog' OO *köbek, *it : OK *it
'sun' OO *gün, güneš : OK *kün.

There are 10 words on Yakhontov's list that are not on Holman's list, but they are not diagnostic.

3.2. Non-basic lexicon

The list of non-basic lexis to reconstruct is quite long. In this article, a few examples show the words characteristic of Kipchak. The list can be composed of such words as verbs, e.g. **aša-* 'to eat', **oz-* 'to outstrip', **yanč-* 'to break', **yomdar-* 'to collect'; nouns, e.g. **bor* 'wine', **kindik* 'navel', **tïltaɣ* 'cause, reason'; adjectives (often used also as nouns), e.g. **köni* 'straight; upright'; and interrogative pronouns, e.g. **nek* 'why'.

**aša-* 'to eat', used side by side with **ye-* in OT and OU, then in all middle Turkic languages and modern language groups (ED 256); CC *aša-* (G 44), Mağ. 45a *ašaɣïl* 'kul' where it is said to be used beside *ye-* (Toparlı, Çögenli, Yanık 2000: 62); this is clearly a derivative from *aš* 'food'; this word was seemingly little used in MK period, but spread in modern languages, though the east sub-group of Kipchak languages (Kazakh, Karakalpak, Kirghiz) normally use *ğe-*, and *asa-, aša-* has a specific meaning (to eat piece by piece); Noghay *aša-* is a later Crimean Tatar of Tatar loanword, since the word for food in Noghay is *as* (NRS 49, 54);

**bor* 'wine'; it is a Middle Persian loanword in OT (ED 354); in later periods it was used in Karakhanid and MK, i.e. CC (G 65), İdrH 8 'sücü,

şarap', Mağ. 17a *bor* 'al-ḫamr' (Toparlı, Çögenli, Yanık 2000: 21); among the modern languages, it survives in Kar. in the derived forms *borla* and *borlalïq* '1. grapevine. 2. vineyard' (KRPS 131, AJ 91);

**kindik* 'navel'; it exists in all modern Turkic language groups except SW, and is attested since OT (ED 729); Clauson argues that *kindik* is an abstract noun from *kin* 'musk'; at the same time, Clauson shows the other term for 'navel' *köbek* and calls it "a purely Western (?Oğuz) word, not noted before XIII, but prob. older" (ED 688); outside SW, Clauson evidences *köbek* only from Chaghatay; ÈSTJa (III 52) also evidences the word *gö:bek* from some modern NW languages such as Crimean Tatar and Crimean Karaim (where they are certainly Turkish loanwords) and a Tatar dialect, but more significantly, from a range of Mamluk Kipchak languages, i.e. Mağ. *köbek*, though without any indication that this is a Trkm. equivalent of Kipchak *kindik* (wrong vocalisation *kandk*), see in 22a (Toparlı, Çögenli, Yanık 2000: 28);

**köni* 'straight; upright', attested in OU (ED 726) and in many Middle Turkic languages; in MK, *köni, könü* is attested in CC (G 151), Gül. *köni* (B 296), Irš. 44a *köni* (Toparlı 1992: 171); according to Clauson, it is derived from *kön-* (many meanings), similarly in ÈSTJa III 75; also attested in Armeno-Kipchak *k'onu* (Tryjarski 1968: 408) and Kar. *köni könü könnü* (KRPS 338, AJ 217);

**nek* 'why', among MK languages, it is evidenced only in CC (G 171), but there is also Armeno-Kipchak *nėk'* (Tryjarski 1969: 547); then we have evidence from CKar., TKar. (KRPS 419, 421, AJ 251) and Urum *nek* (Garkavets 2000: 314); Grønbech assumes the origin from *ne* + *ök*, but it can be a syncopated form of *nelük* 'why', or *nege* 'id', as OT *ičegü* 'intestines, entrails' > *ičeg* > *iček*.

**oz-* 'to outstrip, to surpass', evidenced in OT and OU as well as in later historical languages; although Clauson says that this verb survives in all modern language groups (ED 279), among the SW languages it was rec-

orded only in dialect Turkish (DS IX 3305–3306) and Turkmen (TRS 484); in contrast, *oz-* is found in all modern Kipchak languages and many MK, i.e. Bul. V (Zajączkowski 1938: 63), CC (G 181), Gül. (B 334), Idr. (Ö 591–592), Irš. 449a (T 439), Qav. 79b *ozyïl* 'isbaq' or 'usbuq' (Toparlı, Çögenli, Yanık 1999: 89, 128 mistakenly read it *üz-*; as is known, the verb *üz-* does not have such a meaning, it means 'to tear, to pull to pieces'), Mağ. 37a 'usbuq' (Toparlı, Çögenli, Yanık 2000: 51, 131), Mun. Ö (114), Tuh. FZ *ozdï* (139), Tuh. A mistakenly *uzdï* (38); since this monosyllabic verb must be the root, the dictionaries do not seek etymology, only ÈSTJa (I 425) provides a hypothetic Mongol parallel following Poppe;

**tïltay ~ tïlday* 'cause, reason', evidenced in OT, Karakhanid, but in later periods only in Middle Kipchak (ED 494); the only evidence from MK is Mağ. 39b *dïldayïl* 'aḥtiğğ min al-iḥtiğāğ' and *dïlday* 'al-ḥuğğa' (Toparlı, Çögenli, Yanık 2000: 55, 102); quite interestingly, this word was replaced with the reborrowing from Mongolian already in the 13th–14th centuries in the forms *šïltay* > *sïltaw* etc., e.g. CC *sïltov* (G 228) and the verbs like *sïlta-*, e.g. Tuh. 9b *sïltadï* 'taḥağğağa' (A 237, FZ 373); it is these forms that are used in modern languages, e.g. TKar, HKar. *sïltav* (KRPS 491–492), CKar. *sïltïv* (AJ 350), Kaz. *sïltav* (KRS 750–751); as Clauson has shown, *tïltay ~ tïlday* is a derivative from *tïlta- ~ tïlda-* 'to make excuses, to seek pretexts', though the principal meaning of *tïlda-* should be 'to use the tongue', for it is derived from *tïl* 'tongue'; if this etymology is correct, we have to postulate the back variant **tïl* for at least early Old Kipchak (see 3.1, above), as was the case with OT, OU and partly Karakhanid (ED 489);

**yanč-* 'to crush, to trample', evidenced in OU and the subsequent historical languages (ED 944); while in the SW languages only some dialectal forms are attested (for the forms and languages see ÈSTJa IV 184–185), the NW languages use this verb more extensively; the verb *yanč-* is attested in the following MK documents, CC (G 113), Gül. 'to bore, to pierce' (B 430), Tuh. 16a *yanšïdï* 'darasa' (A reads it *yenşidi* 31, but in the index provides

the back form *yanşı-* 278; FZ 310 *yenši-*); it is an open question if this verb is primary or derived with the suffix *-č-* as Ramstedt (1949: 11) argued, see also Räsänen (1969: 185) and ÈSTJa IV 186; according to Ramstedt, the OT form was *jań- ~ janč-*, similar to *sań- ~ sanč-*, but in fact it is not attested as such and Räsänen's examples, in particular Khakas *nas-*, point in fact to **ÿač- < *ÿaŋč-*;

**yomdar-* 'to collect'; evidenced in Kharakhanid (ED 935), it is a rare, scantily attested verb; among the MK languages, it is found in CC (G 126) and in 16th-century TKar., copied 1686 (Jankowski 2014: 52); according to Clauson, this verb is a causative form of *yomït-* 'to come together, to assemble', which is probably derived from an unattested **yom-*, since *yomyï* 'collected' and there is also the passive form *yomul-* (Jankowski 2014: 52); in fact, it is not needed to derive *yomdar-* from *yomït-*, for the sufix *-DAr-* also exists, see Zajączkowski (1932: 123–124) who among others discusses the Karaim verb *yomdar-*.

There are also some *hapax legomena*, e.g. *čalït* 'diligent', *oy* 'grey', *qar* 'defect'. Especially in Mağ., there is a number of names of fruits and vegetables which do not occur in other document and languages, e.g. *čomar* 'carrot' or *elestek* 'cucumber'.

4. Conclusion

Old Kipchak and Oghuz were certainly different, but similar languages. Although both Middle Kipchak and Middle Oghuz languages are well documented, phonetic features are not well distinguishable in written sources. Lexical reconstruction based on distinctive lexical units has a weakness, i.e. lack of a word in either language may be caused by insufficient documentation. The words pertaining to the basic lexicon mostly differ only phonetically. Therefore, among such words as the numerals from 'one' to 'five'

and some higher, which are regarded as basic (Swadesh 1972: 183), there are only phonetic differences, e.g. OK *eki* : OO *iki* 'two', OK *toquz* : OO *dokuz* 'nine'. The same is true of personal pronouns. Lexical differences in non-basic lexicon are observed in nouns, verbs, adjectives and pronouns, as well as in all grammatical word classes, such as conjunctions, postpositions and modal words.

To reconstruct a more extensive lexicon it is necessary to establish the lexical inventory of both basic and non-basic words. Special attention should be drawn to the words evidenced only in one document, e.g. *oy* 'grey'.

Abbreviations

Kipchak documents

Bul. N – *Kitāb bulġat al-muštāq fī luġat at-Turk wa al-Qifǧāq*, see Zajączkowski 1938

Bul. V – *Kitāb bulġat al-muštāq fī luġat at-Turk wa al-Qifǧāq*, see Zajączkowski

CC. – Codex Cumanicus, see Grønbech 1942, abbreviated G

Dur. – *Kitāb al-durra al-muḍiyya fī luġat at-turkiyya*, see Toparlı 2003

Gül. – *Gülistān bi't-Turkī*, see Bodrogligeti 1969, referred to as B; for the facsimile see Uzluk 1954

Idr. – Abū Ḥayyān's Kitāb al-idrāk li lisān al-Atrāk, referred to as Ö

Idr. H – words on the margins of Al-idrāk excerpted by Rifat Kilisli, glossed and published by İzbudak, see İzbudak 1936

Ilm. – *Kitāb fī 'ilm an-nuššāb*, see Öztopçu 2002

Irš. – *Iršād al-mulūk wa as-salāṭīn*, see Toparlı 1992

Mağ. – *Kitāb mağmūᶜ tarǧumān turkī, ᶜağamī wa muġālī*, see Toparlı, Çögenli, Yanık 2000

Mun. Ö – Munyat al-ġuzāt, Öztopçu 1989

Qav. – *Al-qawānīn al-kulliyya li-ḍabṭ al-luġa at-turkiyya*, see Toparlı, Çögenli, Yanık 1999

Tuh. A – *At-tuḥfa az-zakiyya fī luġat at-turkiyya*, see Atalay 1945

Tuh. FZ – *At-tuḥfa az-zakiyya fī luġat at-turkiyya*, see Fazylov, Zijaeva 1978

Other abbreviations

AJ – Aqtay, Jankowski 2015

CKar. – Crimean Karaim

DLT – Dankoff, Kelly 1982–1985

DS – *Derleme Sözlüğü* 1963–1979

ED – Clauson 1972
ÈSTJa – Sevortjan et al. 1974–2003
HKar. – Halicz Karaim
Kaz. – Kazakh
KRPS – Baskakov, Zajončkovskij, Šapšal 1974
KRS – Syzdykova, Xusain 2001
MK – Middle Kipchak
NRS – Baskakov 1963
NW North-West
OK – Old Kipchak
OO – Old Oghuz
OT – Old Turkic
OU – Old Uyghur
SW South-West
TKar. – Troki Karaim
Trkm. – Turkmen
TRS – Baskakov, Xamzaev 1956
TS = *Tarama Sözlüğü* 1963–1972
WOT – West Old Turkic

References

Alimov, Risbek. 2014. Divanu Lûgati't-Türk'e Göre Kıpçakların Dili: Hayati Develi (et al. eds.) *Maḥmūd al-Kāšġarī'nin 1000. Doğum Yıldönümü Dolayısıyla Uluslararası Dīvān Luġāti't-Turk Sempozyumu. The Dīvān Luġāti't-Turk International Symposium. In Commemoration of Maḥmūd al-Kāšγarī's 1000th Birthday.* İstanbul: Eren, 147–158.

Aqtay, Gulayhan, Henryk Jankowski. 2015. *A Crimean Karaim-English dictionary. 10 000 entries.* Poznań: Department of Asian Studies [= Prace Karaimoznawcze 2].

Atalay, Besim. 1945. *Ettuhfet-üz-Zekiyye fil-Lûgat-it-Türkiyye.* İstanbul: Türk Dil Kurumu [facsimile of low quality, partially hardly readable; a better quality facsimile was published in Halasi Kun 1942].

Baskakov 1963 = Баскаков, Н. А. (ed.). 1963. *Ногайско-русский словарь. Около 15 000 слов. С приложением грамматического очерка ногайского языка, состовленного Н. А. Баскаковым.* Москва: Государственное издательство иностранных и национальных словарей.

Baskakov, Xamzaev 1956 = Баскаков, Н. А., М. Я. Хамзаев. (ed.) 1956. *Туркменско-русский словарь. 50 000 слов.* Москва: Государственное издательство иностранных и национальных словарей.

Baskakov, Zajončkovskij, Šapšal 1974 = Баскаков, Н. А., А. Зайончковский, С. М. Шапшал (eds). 1974. *Карашмско-русско-польский словарь.* Москва: Русский Язык.

Berta, Árpád. 1998. Middle Kipchak: Lars Johanson, Éva Á. Csató (eds). 1998. *The Turkic Languages.* London, New York: Routledge, 158–165.

Bodrogligeti, A. 1969. *A fourteenth century Turkic tranlation of Sa'dī's Gulistān (Sayf-i Sarāyī's Gulistān bi't-turkī).* Budapest: Akadémiai Kiadó [2nd edition in Bloomington at Indiana University Press; no facsimile; facsimile published in Uzluk 1954].

Brands, Horst Wilfrid. 1973. *Studien zum Wortbestand der Türksprachen. Lexikalische Differenzierung, Semasiologie, Sprachgeschichte.* Leiden: E. J Brill.

Clauson, Sir Gerard. 1972. An *etymological dictionary of Pre-Thirteenth-Century Turkish.* London: Oxford UP (at the Clarendon Press).

Dankoff, Robert, James Kelly. 1982–1985. *Maḥmūd al-Kāšγarī. Compendium of the Turkic Dialects (Dīwān Luγāt at-Turk). Edited and translated with introduction and indices by Robert Dankoff in collaboration with James Kelly.* Harvard: Harvard University Printing Office [part I – 1982, part II – 1984, part III – 1985; = Turkish Sources VII].

Derleme Sözlüğü I-XI. 1963–1979. Ankara: Türk Dil Kurumu.

Drimba, Vladimir. 2000. *Codex Comanicus. Édition diplomatique avec fac-similés.* Bucarest: Editura Enciclopedică.

Eckmann, János. 1964. Memlûk Kıpçakçasının Oğuzcalaşmasına Dair. *Türk Dili Araştırmaları Yıllığı Belleten* 1964 [1965], 35–41.

Eckmann, János. 1982–1983. Memlûk Kıpçak Edebiyatı. *Türk Dili Araştırmaları Yıllığı Belleten* 1982–1983 [1986], 85–99.

Fazylov, Zijaeva 1978 = Фазылов, Э. И., М. Т. Зияева. 1978. *Изысканный дар тюркскому языку (грамматический трактат XIV в. на арабском языку).* Ташкент: Фан.

Gabain, Annemerie von. 1959. Die Sprache des Codex Cumanicus: Jean Deny, Kaare Grønbech, Helmut Scheel and Zeki Veledi Togan (eds.). *Philologiae Turcicae Fundamenta, Vol. 1*. Wiesbaden: Franz Steiner, 46–73.

Golden, Peter B. 1992. *An introduction to the history of the Turkic peoples.* Wiesbaden: O. Harrassowitz.

Golden, Peter B. 2014. Qıpčaq: Nurettin Demir, Birsel Karakoç, Astrid Menz (eds). *Turcology and linguistics. Éva Ágnes Csató Festschrift.* Ankara: Hacettepe Üniversitesi, 183–202.

Grønbech, K. 1942. *Komanisches Wörterbuch.* København: Einar Munksgaard.

Halasi Kun, T. 1942. *La langue des kiptchaks d'après un manuscrit arabe d'Istanboul.* Budapest: Société Kőrösi Csoma [Bibliotheca Orientalis Hungarica IV].

Holman, Eric. W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica* 42, 2, 331–354.

İnan, Abdülkadir. 1953. XIII.–XV. Yüzyıllarda Mısır'da Oğuz-Türkmen ve Kıpçak Lehçeleri ve "Halis Türkçe". *Türk Dili Araştırmaları Yıllığı Belleten* 1, 53–71.

İzbudak, Velet. 1936. *El-idrâk Haşiyesi.* İstanbul: Devlet Basımevi.

Jankowski, Henryk. 2014. Two Karaim religious poems by Isaac ben Abraham Troki. *Karaite Archives* 2, 35–57.

Kmietowicz, Anna, Franciszek Kmietowicz, Tadeusz Lewicki. 1985. *Źródła arabskie do dziejów słowiańszczyzny. Tom III.* Wrocław, Warszawa etc.: Zakład Narodowy im. Ossolińskich.

Kuryžanov 1970 = Курышжанов, А. К. 1970. *Исследование по лескике старокыпчакского письменного памятника XIII в. – "Тюркско-арабского аловаря".* Алма-Ата: Наука.

Kuun, Géza comes. 1880. *Codex cumanicus. Bibliothecae ad Templum Divi Marci Venetiarum.* Budapestini: Editio Scient. Academiae Hung. [reprinted end

prefaced by Louis Ligeti in 1981 in: Budapest Oriental Reprints Series B 1, eds E. Schütz, É. Apor].

Mándoky Kongur István. 1993. A *kun nyelv magyarországi emlékei.* Karcag [edited by Júlia Bartha Ökrösné].

Marquart, J. 1914. Über das Volkstum der Komanen: W. Bang, J. Marquart. *Osttürkische Dialektstudien* [= *Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Philologisch-historische Klasse, Neue Folge* 13,1], 25–238.

Menges, Karl Heinrich. 1951. The Oriental elements in the vocabulary of the oldest Russian epos, the Igor' Tale, Slovo O Pŭlku Igorevě. New York: Linguistic Circle of New York [= Supplement to Word, vol. 7, Monograph no. 1].

Musaev, Čečenov 2002 = Мусаев, К. М., А. А. Чеченов. 2002. Кыпчакская группа: Э. Р. Тенишев (ed). 2002. *Сравнительно-историческая грамматика тюркских языков. Региональные реконструкции.* Москва: Наука, 216–338.

Öztopçu, Kurtuluş. 1989. *Munyat'ul-Ghuzāt. A 14th-Century Mamluk-Kipchak Military Treatise.* Washington: Harvard University Office of the University Publisher [Turkish Sources XI].

Öztopçu, Kurtuluş. 2002. *A 14th-Century Archery Treatise in Mamluk-Kipchak.* كتاب في علم النشاب *Kitāb fī 'Ilm An-Nuşşāb.* İstanbul [=Türk Dilleri Araştırmaları Dizisi 34].

Özyetgin, A. Melek. 2001. *Ebū Ḥayyān. Kitābu'l-İdrāk li Lisāni'l-Etrāk. Fül: Tarihi-Karşılaştırmalı Bir Gramer ve Sözlük Denemesi.* Ankara: Köksav [= Tengrim Türklük Bilgisi Araştırmaları Dizisi 3].

Pritsak, Omeljan. 1959. Das Kiptschakische: Jean Deny, Kaare Grønbech, Helmut Scheel and Zeki Veledi Togan (eds.). 1959. *Philologiae Turcicae Fundamenta, Vol. 1.* Wiesbaden: Franz Steiner, 74–87.

Ramstedt, G. J. 1949. *Studies in Korean Etymology.* Helsinki [= Mémoires de la Société Finno-Ougrienne 95].

Ramstedt, G. J. 1957. *Einführung in die altaischen Sprachwissenschaft. I Lautlehre. Bearbeitet und herausgegeben von Pentti Aalto.* Helsinki [= Mémoires de la Société Finno-Ougrienne 104: 1].

Räsänen, Martti. 1969. *Versuch eines etymologischen Wörterbuchs der Türksprachen.* Helsinki: Suomalais-Ugrilainen Seura.

Róna-Tas, András. 1991. *An introduction to Turkology.* Szeged: József Attila Tudományegyetem.

Róna-Tas, András. 1998. The reconstruction of Proto-Turkic and the genetic Qquestion: Lars Johanson, Éva Á. Csató (eds). 1998. *The Turkic Languages.* London, New York: Routledge, 67–80.

Róna-Tas, András, Árpád, Berta, with the assistance of László Károly. 2011. *West Old Turkic. Turkic loanwords in Hungarian. Part I. Introduction, Lexicon »A–K«, Part II. Lexicon »L–Z«, Conclusions, Apparatus.* Wiesbaden: Otto Harrassowitz.

Sevortjan et al. 1974–2003 = *Этимологический словарь тюркских языков. т. I* (1974); *т. II* (1978); *т. III* (1980). Э. В. Севортян; *т. IV* (1989). Э. В. Севортян, Л. С. Левитская, А. В. Дыбо. Москва: Наука; *т. V* (1997). Л. С. Левитская, А. В. Дыбо, В. И. Рассадин. Москва. Языки Русской Культуры; *т. VI* (2000). Л. С. Левитская, А. В. Дыбо, В. И. Рассадин. Москва: Издательство Индрик; *т. VII* (2003). Л. С. Левитская, Г. Ф. Благова, А. В. Дыбо, Д. М. Насилов. Москва: Восточная Литература РАН.

Swadesh, Morris. 1972. What is glottochronology?: Morris Swadesh. *The origin and diversification of language.* London: Routledge & Kegan Paul [edited by Joel Sherzer], pp. 271–284 [the article translated from Spanish by Joel Sherzer].

Syzdykova, Xusain 2001 = Сыздыкова, Р. Ғ., К. Ш. Хусаин (ed.). 2001. *Казахско-русский словарь. Около 50000 слов.* Алматы: Дайк-Пресс (Сыздықова, Р. Ғ., К. Ш. Хұсайын. 2001. *Қазақша-орысша сөздік. 50 мыңға жуық сөз.* Алматы: Дайк-Пресс).

Tarama Sözlüğü. Vol. I-VI. 1963–1972. Ankara: Türk Dil Kurumu.

Tenišev 1997 = Тенишев, Э. Р. (ed). 1997. *Сравнительно-историческая грамматика тюркских языков. Лексика.* Москва: Наука.

Toparlı, Recep. 1992. *İrşâdü'l-Mülûk ve's-selâtîn.* Ankara: Türk Dil Kurumu [no facsimile].

Toparlı, Recep. 2003. *Ed-Dürretü'l-Mudiyye fi'-l-Lügati't-Türkiyye.* Ankara: Türk Dil Kurumu.

Toparlı, Recep, M. Sadi Çögenli, Nevzat H. Yanık. 1999. *El-Kavânînü'l-Külliye li-Zabtil'l-Lügati'l-Türkiyye.* Ankara: Türk Dil Kurumu.

Toparlı, Recep, Sadi Çögenli, Nevzat H. Yanık. 2000. *Kitâb-ı Mecmû-ı Tercümân-ı Türkî ve Acemî ve Mugalî.* Ankara: Türk Dil Kurumu.

Toparlı, Recep, Hanifi Vural, Recep Karaatlı. 2003. *Kıpçak Türkçesi Sözlüğü.* Ankara: Türk Dil Kurumu.

Tryjarski, Edward. 1968–1972. *Dictionnaire arméno-kiptchak d'après trois manuscrits de collections viennoises. Tome I* (fascicule I et II – 1968, fascicule III – 1969, fascicule IV – 1972). Warszawa: Zakład Orientalistyki Polskiej Akademii Nauk.

Uzluk, Feridun Nafiz. 1954. *Seyfî Serâyî Gülistan Tercümesi.* Ankara: Türk Tarih Kurumu.

Zajączkowski, Ananiasz. 1932 *Sufiksy imienne i czasownikowe w języku zachodniokaraimskim (Przyczynek do morfologji języków tureckich). Les suffixes nominaux et verbaux dans la langue des Karaïms occidentaux (Contribution à la morphologie des languaes turques). Avec résumé français.* Kraków: Polska Akademja Umiejętności.

Zajączkowski, Ananiasz. 1938. *Manuel arabe de la langue des turcs et des kiptchaks (Époque de l'État Mamelouk). Arabski podręcznik języka Turków i Kipczaków z epoki Państwa Mameluckiego (Introduction, vocabulaire turc–polonais–français, texte). Avec 4 planches. [1-ère partie. Le nom.].* Warszawa: Towarzystwo Naukowe Warszawskie.

Zajączkowski, Ananiasz. 1949. *Związki językowe połowiecko-słowiańskie.* Wrocław: Wrocławskie Towarzystwo Naukowe.

Zajączkowski, Ananiasz. 1954. *Słownik arabsko–kipczacki z okresu Państwa Mameluckiego. Bulġat al-muštāq fi luġat at-Turk wa-l-Qifǰāq. Część II. Verba.* Warszawa: Państwowe Wydawnictwo Naukowe [= Prace Orientalistyczne I].

ABSTRACT

This paper attempts to work out some principles of the establishment of those Old Kipchak words which are distinctive from other Old Turkic languages. Old Kipchak is a language that was spoken in two periods, one Early Old Kipchak from the late-8th century when it first appeared to the mid-11th century, the other Late Old Kipchak from the second half of the 11th century to the Mongol invasion in the 1230s. There is only scanty documentation of Old Kipchak from its late period by Maḥmūd al-Kašġārī, often ambiguous, the remaining being single glosses in Old Russian, Arabic and other documents. Phonological criteria are of limited applicability because of the weakness of the phonetic value of written texts. Despite these limitations, there is some lexical material with which we can start the reconstruction. The reconstructed lexicon should include both basic and non-basic lexicon. Since common Old Turkic languages, expect for Old Bulgar, were similar, basic lexicon will be similar to East Old Turkic, which is well documented. Non-basic lexicon is composed of those lexical units which are different from the most similar Old Oghuz and those which are similar or identical. This paper focuses on the words which are unknown in Oghuz languages. One can assume that the difference has arisen by either real lack of the selected words in one group of languages or their disappearance from use.