Khitan: Understanding the Language Behind the Scripts

Juha A. Janhunen
University of Helsinki, Finland

Khitan was the dynastic language of the Liao empire in Manchuria and Northern China (907-1115). Although today extinct, samples of Khitan are preserved in two native scripts, known as the Khitan Large Script and the Khitan Small Script. Both scripts may be classified as "Sinitic" or "Sinoform" in the typological sense, though only the Large Script has a direct connection with the Chinese script. Recent progress in the decipherment of, in particular, the Khitan Small Script allows the lexicon and grammar of the Khitan language to be assessed in much more detail than before. Khitan may be defined as a Para-Mongolic language, meaning that it represents a branch related to, but collateral with, the extant and historically known Mongolic languages. The present paper examines the genetic position of the Khitan language with regard to Mongolic with the help of the methods of comparative linguistics, as applied to the deciphered Khitan language material.

Keywords: Sinitic scripts, Khitan Small Script, Khitan language, Para-Mongolic, reconstruction, decipherment

In the global context of ancient scripts and their decipherment it is not widely known that some of the last great challenges posed by unknown systems of writing are those connected with the scripts of peripheral mediaeval China. Three non-Chinese ethnic groups—the Khitan, the Jurchen, and the Tangut, corresponding to the "alien" dynasties of Liao 遼 (907-1125), Jin 金 (1115-1234), and Xixia 西夏 (1038-1227), respectively—created scripts and literary languages of their own. After centuries of oblivion, samples of these scripts emerged in the late 19th and early 20th centuries and became objects of intensive study. Even so, we cannot say

SCRIPTA, Volume 4 (October 2012): 107-132
© 2012 The Hunmin jeongeum Society

that any of these scripts has been fully deciphered. The problems vary from script to script, however, and they are also connected with the level of knowledge we have of the underlying languages—Khitan, Jurchen, and Tangut. It happens that of these three languages, only Jurchen survived till later times, becoming the direct ancestor of Manchu, while both Khitan and Tangut became extinct soon after the fall of the political states that had used them as their dynastic languages.[1]

1. The Khitan Scripts

As far as Khitan is concerned, the situation is complicated by the circumstance that it was written in two different scripts, today known as the Khitan Large Script (大字 *Dazi*) and the Khitan Small Script (小字 *Xiaozi*). Like the Tangut Script and the Jurchen Script, the two Khitan scripts have a Chinese "appearance," meaning that they are composed of the same type of primary and secondary elements (strokes, characters), presupposing the same type of writing instruments (brush, ink), as the Chinese script. We may therefore identify all these scripts as "Sinitic" or "Sinoform" in the typological and aesthetic sense. None of them is, however, identical with the Chinese script, and only two scripts, the Khitan Large Script and the Jurchen Script, are materially "related" to the Chinese script, meaning that they may be viewed as ancient regional derivatives of the latter, a situation which also allows them to be identified as "Sinographic" writing systems, and their characters as "Sinograms." The Khitan Small Script and the Tangut Script, on the other hand, have no material relationship with the Chinese script, which suggests that their origination may have involved a factor of conscious "invention."[2]

We know today that the Khitan Large Script and the Jurchen Script are very closely related with each other, both probably representing a

[1] For a more comprehensive survey of the chronological context of the scripts of mediaeval peripheral China, cf. Janhunen (1994).

[2] The present paper will not deal in any more detail with the Tangut script. However, from the point of view of graphemic analysis, this script offers a particularly difficult challenge, since nothing is known of the principles of how it was designed. For this reason, the assumptions conventionally made about the identity of the Tangut language are also detached from the graphic reality and cannot be considered as confirmed, as was already pointed out by Kwanten (1984, 1988).

continuation of a local Manchurian variety of the Chinese script. It is possible, though not verified, that this same variety, or an earlier form of it, was used already before the Liao-Jin period in other political contexts, including, perhaps, Beiwei 北魏 (386-534) and/or Balhae 渤海 (698-926). It is notable that the Khitan Large Script incorporates many Chinese characters, especially simple ones, in unchanged form, while in the Jurchen Script such characters are normally distinguished from their Chinese counterparts by using diacritics. We also know that both the Khitan and the Jurchen used this script as a mixed logo-syllabic system, with some characters functioning as logograms (with Khitan and Jurchen pronunciations) and others as syllabograms (for fixed sequences of sounds, not necessarily corresponding to phonetic syllables). In general, the number of separate characters in the Khitan Large and Jurchen Scripts is considerably smaller than in the Chinese script, and the forms of the characters are less complicated, with no functionally relevant "radical" components.

Unfortunately, the Khitan Large Script is still very imperfectly deciphered, with only a small proportion of the characters identified with meanings (logograms) and/or sounds (syllabograms). One important direction of research that has not yet been properly initiated is the systematic comparison of the Khitan Large Script characters with those of the Jurchen Script. Due to the fact that the Jurchen Script survived relatively long (till the 16th century) and was used to write a well-known language (Jurchen-Manchu), our knowledge of the Jurchen Script is rather close to "decipherment" in the sense that most of its elements have a known linguistic correlate, even if there still remain many open questions.[3] On the other hand, the corpus of the Khitan Large Script is much more extensive, and in a much better state of preservation, than that of the Jurchen Script. Therefore, the Khitan Large Script is potentially a crucial source of information also for Jurchen studies. At the present stage, however, its potential cannot yet be exploited.

This leaves us only the Khitan Small Script corpus as the most important currently available direct source on the Khitan language. Since several decades, it has been known that this is a basically syllabic script,

[3] The principal works on the Jurchen Script are Grube (1896), Kiyose (1978), Kane (1989), Jin Qicong (1984), and Pevnov (2004), as well as, most recently, Golovachev & Ivliev & Pevnov & Rykin (2011). Of these, only Kane comments also on the Khitan Large Script.

using a limited number of graphically simple syllabic signs (syllabograms), known as the Khitan Small Script "characters." What makes the graphic image visually complicated is the convention of accumulating the syllabic signs into "blocks" corresponding to linguistic words, a convention which, incidentally, has a close analogy in the Korean Hangeul.[4] Since a Khitan word (with derivational and inflectional suffixes) can consist of up to eight "syllables," a block can have a graphic appearance (stroke composition) even more complex than a Chinese character—a circumstance that caused confusion during the early stages of Khitan studies, when it was not yet clear how the "Small" and "Large" scripts should be properly distinguished.[5]

The decipherment of the Khitan Small Script has involved both slow progress and rapid steps forward starting with the early 1920s.[6] A definitive breakthrough, which for the first time yielded concrete "readings" of Khitan words, was achieved by a team active in China in the '70s and '80s.[7] Since then, the most important development has been the rapid increase of the corpus, conditioned by the discovery of new texts. The currently known corpus comprises 34 published and a few unpublished epigraphic texts plus a diffuse selection of minor materials. The standard of "publication" varies, however, and not all texts have been made available in a form that can be used for textological and linguistic analysis. Even so, and in spite of the fact that the epigraphic texts tend to be stereotypic in form and content, the corpus is large enough to allow relatively detailed conclusions to be made about the Khitan language and the properties of the Small Script.

The phonetic value of the Khitan Small Script syllabic signs varies from

---

[4] The difference between the two scripts is, of course, that the blocks in Hangeul are composed of phonograms. It is not known whether the Khitan Small Script can have inspired the creators of the Hangeul, for ultimately the blocks in both the Khitan Small Script and the Hangeul must have been modelled along the principle of Chinese characters.

[5] The distinction between the two Khitan scripts was first correctly made and communicated internationally by Toyoda (1964).

[6] A large part of the research on the Khitan Small Script has been communicated in scattered papers and a few monographs in Chinese. An important critical summary of all this work was made available for the international readership by Kane (2009).

[7] Chingeltei & Liu Fengzhu & Chen Naixiong & Yu Baolin & and Xing Fuli (1985). For a more detailed history of Khitan Small Script studies, see Wu Yingzhe & Janhunen (2010: 20-25).

one segment (consonant or vowel) to sequences of up to three segments. Some characters seem to be used also for longer words as what may be actual logograms. There are, of course, also undeciphered characters. Of the currently known 459 different characters, "only" 314 have been deciphered in one way or another.[8] These include, however, a large proportion of the most frequent characters, while many of the undeciphered items are rare and may even involve mistakes. This means that even if the decipherment is not fully completed, it is already possible to render long sequences of continuous Khitan text a more or less consistent phonetic and/or semantic interpretation, with only occasional lacunae. In this situation, attention may gradually be turned away from the script to the language behind the script.

2. The Khitan language

The Khitan, or *Qidan* 契丹, were the dominant ethnic group in the Liaoxi 遼西 region in the period between the Beiwei and the historical Mongols. The Khitan survived as an ethnic group after the rise of the Jurchen, and written documents in the Khitan language continued to be produced almost till the end of the Jin period. It is well known from Chinese historical sources that Khitan was a distinct language, spoken by virtually all ethnic Khitan, a population whose size may have reached a million people during the height of the Liao dynasty. Khitan was not the only language spoken in the Liao state, however, for other languages, including Chinese and Jurchen, were also widely used especially in the areas conquered by the Khitan during the process of state formation. It is less clear how uniform Khitan was internally. The Khitan were composed of tribes, some of which played a more important political role in the Liao state than others, and it is quite possible that there were tribal forms of speech that deviated from the mainstream language.[9]

Apart from texts in the two Khitan scripts, there are two other types of sources on the Khitan language. These are, first, the occasional samples of Khitan, mainly lexical items, but also phrases and poems, preserved

[8] An up-to-date list of the Khitan Small Script characters is given in Wu Yingzhe & Janhunen (2010: 259-272, cf. also the discussion ibid. 35-48).

[9] On the tribes, as well as on the general social history and historical demography of the Khitan, see Wittfogel & Fêng (1949).

in Chinese transcription and glossing, and, second, the Khitan loanwords transmitted into neighbouring languages, especially Jurchen. As far as can be seen, these sources represent a language identical with that of the Khitan text corpus. The Chinese data are, however, often frustratingly difficult to interpret with any certainty due to the chronological, phonological, and semantic inaccuracies as well as outright errors contained in them.[10] Loanword research could potentially yield much more informative results, for it may be taken for certain that Jurchen, as a subordinate language of the Liao state, received a considerable number of Khitan loanwords.[11] It is also likely that the two languages were typologically similar, which means that we may occasionally use information from Jurchen to approach the structure of Khitan.

The fact that Khitan is, or was, typologically speaking, an "Altaic" language, was known already before the original sources in the language could be "read."[12] This is now confirmed by the actual "readings" of the Khitan Small Script texts. We know, for instance, that Khitan had an elaborate system of nominal and verbal inflectional suffixes, most of which can be phonologically approximated.[13] The syntax and morphosyntax, including the basic word order (SOV) and the various types of subordinated (converbial) and embedded (participial) sentence constructions, is also in full accordance with that attested in the other historical and modern "Altaic" languages in the region, including both Jurchen-Manchu and Mongolian. There are, however, some unexpected features, notably, traces of grammatical gender (marked masculine vs. unmarked or generic feminine), a category not typical of "Altaic" typology though, incidentally, present in Middle Mongol.[14]

---

[10] Recent studies of Khitan words in Chinese sources include Sun Bojun & Nie Hongyin (2008) and Talpe (2010). The only linguistically competent work on the subject is Shimunek (2007). Selected details are also discussed by Vovin (2003) and Róna-Tas (2004).

[11] Some pioneering work on the lexical parallels between Khitan and Jurchen has been carried out by Kane (2006).

[12] A "grammar" and a collection of Khitan Small Script "texts" in digital form was compiled by a Russian team in the 1960s and '70s, cf. Starikov *et al.* (1970), Arapov (1982).

[13] So far the only systematic grammatical sketch of Khitan has been compiled by Kane (2009: 131-166), who also gives a vocabulary of 354 items (ibid. 83-130).

[14] The role of grammatical gender in Khitan is still far from completely understood.

It used to be more difficult to determine what genetic group (language family) Khitan represents. "Altaic" typology as such does not indicate whether a language is Turkic, Mongolic, Tungusic, or something else. In the past, when work on Khitan was solely based on the Chinese transcriptions and glosses of Khitan words, it was common to try to identify these words item by item with the different historical and extant "Altaic" languages. The result was ambiguous, since Khitan turned out to contain lexical items that could be identified variously with either Mongolic or Turkic, or also Tungusic.[15] In many cases, parallel identifications were possible, leading to speculations that Khitan might be an "intermediate" idiom between, say, Turkic and Mongolic, or Mongolic and Tungusic. The problem here was that inherited elements were not properly distinguished from borrowed ones.

A more reliable method is to operate with genetically diagnostic words, such as items of basic vocabulary, which are likely to have been inherited rather than borrowed. With this method, it was initially possible to identify the language of the Tabghach, or *Tuoba* 拓跋, the leading ethnic group of the Beiwei state, as basically "Mongolic." Since the Tabghach were the political and, quite possibly, the linguistic ancestors of the Khitan, the Mongolic identification of Khitan gained ground.[16] This identification is now definitively confirmed by the information from the Khitan Small Script texts, which show beyond doubt that Khitan was a language whose basic vocabulary and grammatical resources were related to those attested in the Mongolic languages. It is important to note that this conclusion is exclusive, in the sense that Khitan *is not* an "intermediate" language and *cannot* share elements, except borrowings, with other "Altaic" languages than Mongolic.[17]

The term "Mongolic" needs, however, modification when used about

---

Work on this issue has been carried out by Wu Yingzhe (2005, 2007). In spite of its importance, the issue of grammatical gender is largely unstudied in Middle Mongol, cf. Rybatzki (2003: 75).

[15] Menges (1968) made an effort to identify Tungusic elements in Khitan, though we know today that most of his identifications are wrong. He had to admit himself that the material was not conclusive for the determination of the genetic position of Khitan.

[16] The Mongolic identification of the Tabghach language was made by Ligeti (1971). A similar identification for Khitan is implied by Doerfer (1992, 1993).

[17] Suggestions that violate this principle are occasionally encountered in the works of Khitan specialists not sufficiently familiar with comparative linguistics.

Khitan. All other extant and historical Mongolic languages represent a single genetic node, corresponding to the language of the historical Mongols, and linguistically identifiable as Proto-Mongolic. Khitan did not descend from Proto-Mongolic but was, rather, a sister language of the latter. In other words, Proto-Mongolic and Khitan represented two branches of an even older protolanguage, a certain stage of Pre-Proto-Mongolic that could also be termed "Khitano-Mongolic." Khitan itself may also have had sister languages belonging to the same branch, which, in that case, should be termed "Khitanic." Technically, the most suitable term to describe the position of Khitan (and Khitanic) with regard to Proto-Mongolic is to identify the former as "Para-Mongolic," implying that it is a question of a genetically related, but collateral, branch of the much better known Mongolic language family.[18]

3. The position of Khitan

The availability of direct information on Para-Mongolic in the form of Khitan Small Script texts opens up extremely important new perspectives for studies on the history of the Mongolic language family. Working with the actual Mongolic languages alone, we used to have three kinds of diachronic information: first, the comparative evidence provided by the extant Mongolic languages; second, the information contained in the historical forms of Mongolic, especially Written Mongol and Middle Mongol; and third, the potential conclusions that can be made from the comparative and historical data by the method of internal reconstruction. Only the last type of information allows us to approach the period preceding Proto-Mongolic, which itself is of a rather shallow depth corresponding to no more that 800-900 years.

The question is how far backwards the Khitan data allow us to go in the history of Mongolic. This, on the other hand, depends, on how different Khitan was as compared with Proto-Mongolic. Before the decipherment of the Khitan Small Script there was an opinion that Khitan may have been closely related to Proto-Mongolic, so closely that it was thought that

[18] The term Para-Mongolic was introduced by Janhunen (1995, 2003) and has since found some support in specialist literature, cf. e.g. Kane (2009: x).

Khitan could perhaps be "read" in terms of the lexical and grammatical information we have from Middle Mongol. Today we know that this is not so: Khitan is, in fact, a language rather distantly related to the Proto-Mongolic branch. To get an idea of how distant the relationship could be we may think of the neighbouring Tungusic language family, in which the southern branch, corresponding to Jurchen-Manchu (Jurchenic), is in many ways strikingly different from the northern branch, corresponding to Ewenki-Ewen (Ewenic).[19]

Assuming that the difference between Para-Mongolic and Proto-Mongolic was of the same chronological scope as that between the southern and northern branches of Tungusic, the breakup of the original protolanguage (Proto-Khitano-Mongolic) would have taken place at least several centuries before the emergence of Khitan as a written language. The last possible historical context for the still uniform protolanguage would seem to have been the empire of the Xianbei 鮮卑 (93-234), though the breakup may, of course, have taken place even earlier. A practical consequence of the chronological difference is that Khitan texts, which with one notable exception are not bilingual, are difficult to understand even if the script is no longer a major problem. It is not without reason that the situation has been compared with that of Etruscan: a known script but an unknown language.[20]

Fortunately, however, Khitan—unlike Etruscan—is not completely unknown, for it is still related to the Proto-Mongolic branch, although the relationship is so distant that it does not substantially facilitate the understanding of Khitan texts. The genetic connection is nevertheless evident from the existence of cognate words and shared morphological elements present in both Khitan and Proto-Mongolic. It is very likely that the number of cognates will continue to grow as more linguistic work is done on the Khitan Small Script texts. We should, however, not be too optimistic about the size of the comparative corpus. The total number of Khitan lexical items that can be assessed both phonetically and semantically today is still less than 500, and many of these items do not

[19] The Tungusic family also has two transitional (or mixed) branches, collectively termed "Amur Tungusic." For a more detailed taxonomy of Tungusic, cf. Janhunen (2012).

[20] The Etruscan parallel is mentioned by Kane (2009: x), who also discusses the general challenge posed by languages that can be read without being understood. On the unique bilingual Khitan text, cf., most recently, Vovin (2011).

belong to the basic vocabulary; moreover, they also comprise a number of Chinese borrowings.

As it is, the number of currently known certain or plausible cognates between Khitan and Proto-Mongolic is at the range of a few dozen, that is, about ten per cent of the total known Khitan lexicon.[21] The only way to assess the taxonomic status of Khitan with regard to Proto-Mongolic is to analyze this corpus with the methods of comparative linguistics. It is particularly illustrative to see in what respects Khitan, as compared with Proto-Mongolic, is archaic and in what respects it is innovative. Since Khitan became a written language some 200 years before Middle Mongol, one would expect that it might be in some respects more archaic. The picture is not equivocal, however, for in other respects Khitan is surprisingly "modern," often anticipating innovations that took place in the Proto-Mongolic branch only after the breakup of the latter. This is compatible with the picture provided by the Tungusic family, in which Jurchen-Manchu may also be seen as a particularly innovative branch as compared with Ewenki-Ewen.[22] Moreover, some of the innovations present in Khitan are likely to have been shared with Jurchen on an areal basis.

At this point a word of caution concerning the notation is in place. The language behind the Khitan Small Script can be approached in terms of a Romanized approximation of its phonetic structure. The Romanization is, however, not an exact reconstruction of the segmental properties of the language. This situation is due to two circumstances: on the one hand, we simply do not know the exact segmental counterparts of many Khitan Small Script characters, although we can "approximate" them; on the other, we have to consider the fact that the orthography of Khitan is not always consistent with the actual sequences of sounds. In order to show the difference between orthography and sounds, it is useful to let the Romanizations be accompanied by a more phonemically accurate "reading" (marked †), which, in turn, should be distinguished from reconstructions (marked *).[23]

---

[21] Many of the cognates have been identified over the years by Chinese scholars and are listed by Kane (2009), though his database involves some mistakes and omissions.

[22] The conventional view that Jurchen-Manchu represents an innovative branch of Tungusic against the more conservative Ewenki-Ewen branch has recently been challenged by Alonso de la Fuente (2011). The issue is best seen as unconcluded.

[23] The current system of Romanization for Khitan was introduced by Kane (2008, 2009)

4. Lexical properties

As all Khitan texts of any significant length are memorial in character, containing epitaphs, eulogies, lamentations, and/or historical records, the lexicon used in them is thematically and functionally restricted. This gives the decipherer the advantage of having many parallel repetitions with minor variations, as in genealogical lists, which, in turn, allows the identification of certain crucial lexical items and phrases to be multiply verified. The obvious disadvantage is that large sections of the lexicon never occur in the texts, or occur so rarely that identification is impossible. A potentially important source of lexicon is supplied by the poems which often complement epitaphic texts, but the problem here is that poetic expressions are virtually hopeless to assess without a parallel version in a known language.

In the following, a few lexical spheres relevant to comparative linguistics are examined in some more detail:

Pronouns. For genetic comparisons, personal pronouns would be an ideal source of diagnostic information. Unfortunately, first and second person pronouns do not seem to be attested in the extant Khitan corpus. A possible third person pronoun is 田 *hó*, but this item has no known cognate in Proto-Mongolic. The proximal demonstrative is, however, fully comparable with Proto-Mongolic data. The Khitan shape is 爻 *e* 'this' : 令 *t* 'these,' suggesting the readings †*e* : plural †*e-d*, which correspond to Proto-Mongolic **e-n*=*e* : plural **e-d*=*e*. As may be seen, the Khitan shapes look like more simple and, hence, archaic, as they lack the stem extension *-n* and the postclitic element =*e*, as attested in the Proto-Mongolic cognates.[24]

On the Mongolic side, the proximal pronoun has the oblique stem **e-xü/n-* > **üün-*, which has been compared with the possibly pronominal

---

and developed further by Wu Yingzhe & Janhunen (2010). The phonetic reliability of the Romanized shapes varies, and not all Romanizations may be regarded as fully verified. In the present paper, the Khitan items are quoted both in the original script (linearized) and in Romanization, with a phonetic approximation given when relevant or possible.

[24] All Khitan data are quoted from Kane (2009) and Wu Yingzhe & Janhunen (2010). On the demonstrative pronouns, cf. also Wu Yingzhe (2009, 2011).

Khitan monosyllable 𘰾 *ún* †*un*. This comparison remains unverified, but it would seem to correlate with Khitan 𘰾𘰺 *ún.e* †*une* 'now,' which formally corresponds to the Proto-Mongolic locative case form **e-xü/n-e* 'upon this,' a case form also attested in **e-d/ü-x-e* 'now' (> Modern Mongolian *odoo*). The demonstratives are, however, not necessarily among the most stable elements of a language, which is why there is no need to expect a full correspondence between Khitan and Proto-Mongolic. It is therefore not surprising that the Khitan distal pronoun is 𘰀 *qi* 'that' : plural 𘰀𘰼 *qi.t*, suggesting an item totally different from Mongolic **te-*.[25]

Numerals. One of the most simple demonstrations of the relationship between Khitan and Mongolic is offered by the basic numerals. Most numeral roots in Khitan can be written in two ways: either by a single character, often for cardinals, or in terms of a sequence of characters, often for ordinals. In both cases, it is a question of syllabic signs with a phonetic value, though some of the single characters for numerals still lack a phonetic approximation. From the combined information of the two types of spelling we can establish the following basic numeral stems: 2 𘰇𘰽 *ci.ur-* †*jur-*, 3 𘰋𘰽 *hu.ur-* †*hur-*, 4 𘰼𘰽 *t.ur-* †*dur-*, 5 𘰓 *tau* †*tau*, 7 𘰞𘰾 da_2*.lo-* †*dal-*. The similarity to Mongolic 2 **jir-* (or **jïr-*),[26] 3 **gur-*, 4 **dör-*, 5 **tab-*, 7 **dal-* is unmistakable, and it would be impossible to explain the situation by assuming borrowing. The higher numerals for the powers of ten 100 𘰔 *jau*, 1000 𘰕 *ming*, 10 000 𘰖 *tum* also have Mongolic cognates: **jaxu-*, **mingga-*, and **tüme-*, though they are less diagnostic, since they would have been more liable to be borrowed.

We also know the approximate phonetic shape of the Khitan numeral stem 9 𘰗 *is* †*is-*. This item is often compared with the Modern Mongolian

[25] In this case, the discrepancy has also been explained by drawing the Romanization of the Khitan character into question. Most recently, Wu Yingzhe (2011: 73-76) has suggested the Romanization *te*, congruent with Mongolic **te*, though *mun*, which would correspond to Mongolic **mön* 'this,' has also been proposed. Unfortunately, such reinterpretations are not compatible with the evidence supporting the Romanization *qi*, as established on the basis of another (ethnonymic) attestation by Kane (2009: 73), cf. also Wu Yingzhe & Janhunen (2010: 65-66).

[26] Since the Proto-Mongolic reconstructions are, in principle, well known, the present paper will not discuss them in the detail. It may nevertheless be noted that the numeral root for 2 is harmonically ambiguous: palatal vocalism is suggested by **ji-tüxer* 'second (wife)' and possibly **jirin* 'two (females),' while velar vocalism is suggested by **jïr-a/n* 'sixty' (2x3x10).

numeral 9 *is* ~ *yis*. This comparison is wrong, however, for the Mongolian item derives from **yer-sü/n* (as also in 90 **yer-e/n*), an innovation of the Proto-Mongolic lineage, as is also confirmed by data from the "teens" in Jurchen-Manchu.[27] Another innovation in Mongolic is the numeral 6 **jir-gu-xa/n* (2x3), which must have replaced the original numeral likely to hide behind the Khitan character 𘟥 SIX. Judging by the corresponding "teen" in Jurchen-Manchu, the Khitan item may have been †*nil-*, but this shape cannot yet be confirmed by the Khitan Small Script.

It may be noted further that the Khitan basic numerals seem to lack the stem extension **-pA/n* > **-bA/n* ~ **-xA/n*, which is present in the Mongolic items for the first decade 3 **gur-ba/n*, 4 **dör-be/n*, 7 **dal.u-xa/n*, and others. This element might, then, represent an innovation specific to the Proto-Mongolic lineage, though it is also possible that the Khitan numeral system was secondarily simplified; in fact, we do not know what the exact shapes of the Khitan numerals were in absolute cardinal use. On the other hand, the Khitan numerals 2 †*jur-*, 3 †*hur-*, 4 †*dur-* contain the segment **r*, which most probably is a derivative element. This element is also present in Mongolic, meaning that it must represent an innovation common to both the Proto-Mongolic and the Para-Mongolic lineage.[28]

Terms for the four seasons. Among the surprises offered by the comparative material are the terms for the four seasons, which are virtually identical in Khitan (†) and Proto-Mongolic (*): 𘠽𘟙 *heu.úr* †*haur* = **kabur* 'spring,' 𘍞𘟂 ju_3*.un* †*jun* = **jun* 'summer,' 𘟪𘍞𘟙 *n.am.úr* †*namur* = **namur* 'autumn,' 𘟃𘠋 *u.ul* †*uul* = **öbül* (< **ebül*) 'winter.' Since the possibility of a wholesale borrowing is extremely unlikely, it must be a question of cognates. Moreover, the Khitan items, when not fully identical with their

[27] On the teens in Jurchen-Manchu, cf. Janhunen (1993). Although these items were certainly borrowed from a Mongolic language, we do not know whether the borrowing happened before or after the split of the Proto-Mongolic and Para-Mongolic lineages. The structure of the teens in Jurchen-Manchu differs from that attested in the language of the Khitan texts. On the reconstruction and connections of the Khitan numerals, cf. also Chingeltei (1997).

[28] Using the method of internal reconstruction, the **r* in 3 **gur-* may be compared with the **t* present in Mongolic 30 **guc-i/n* < **gut-ï/n*, though the background of this correspondence is unknown. The case of 4 **dör-* : 40 **döcin* may be more complicated since this basic digit is also attested in some Mongolic languages (Shirongolic) as **der-*, while the corresponding decade is present, as a loanword, in Tungusic in the unexpected shape (*)*deki*(*n*), cf. Doerfer (1985: 78-79).

Proto-Mongolic counterparts, show a somewhat greater degree of phonological innovation. This would seem to confirm that Khitan was generally more innovative, or also more "worn," than Proto-Mongolic.

The terms for the four seasons also provide an important clue to understanding the innovations shared by Proto-Mongolic and Para-Mongolic. The medial nasal **m* in **namur* 'autumn' must represent the same segment as the medial obstruent **b* in **kabur* 'spring.' In other words, it is a question of a suffix, possibly reconstructable as Pre-Proto-Mongolic **-pUr*, which is represented as **-mUr* after a syllable with an nasal onset, and as **-bUr* elsewhere. Possibly, the element **(-)bül* in **öbül* 'winter' is also relevant in this context. In any case, the change **p* > **m*, which may be seen as a case of progressive distant assimilation, is likely to have taken place in the common protolanguage of Proto-Mongolic and Para-Mongolic.[29]

Kinship terms. In view of the numerals and the terms for the four seasons it is curious that there is no overall correspondence in the realm of kinship terminology. Several Khitan kinship terms, including 𘮩 *ai* 'father,' 𘮪 *ia* 'elder brother,' and 𘮫 *au* 'elder sister,' have no obvious cognates on the Mongolic side. In other cases the correspondences are uncertain: 𘮬 *mo* 'woman, mother, wife' (plural 𘮬𘮭 *mo.t*) and 𘮮 *bu* 'grandfather' (in 𘮮𘮩 *bu ai* id.) have been compared with Mongolic **eme* 'woman' resp. **ebügen* 'old man,' but the issue cannot be considered as fully concluded.[30] The comparison between 𘮯𘮰 *b.qo* 'child' (plural 𘮯𘮱𘮲 *b.hu.án*) and Mongolic **baga* 'small' > 'child' is also problematic.

There are, however, also good and non-trivial correspondences. These include 𘮳𘮲 *na.ha* †*naha* (plural 𘮳𘮲𘮴𘮵 *na.ha.án.er*) 'maternal uncle' = **naga-cu id.*, 𘮶𘮷 *k.iú* ~ 𘮸𘮷 *x.iú* †*keu* 'younger sister' = **keü* 'child,' and very probably 𘮺 *deu* †*deu* (genitive 𘮺𘮻 *deu.un*) 'younger brother' =

[29] The reconstruction of the original consonant behind the variation **b* ~ **m* is not a trivial task, of course, and it is possible that **b* would be a better choice than **p*. In other items, the variation also involves **x*, as in the numerals, in which **m* is present in Mongolic 8 **nai-ma/n*. For the time being we do not know what the phonetic shape of the corresponding Khitan numeral was. On the Khitan terms for the four seasons, cf. also Toyoda (1998).

[30] Wu Yingzhe (2009) argues, certainly correctly, that in some cases a Khitan character can imply a "hidden" initial vowel. If the Khitan phonemic shapes were something like †*eme* resp. †*ebu* the cognateship with Mongolic would be in no doubt.

**de(x)üi* id.[31] In two items, Khitan would seem to have a monosyllabic root (CV) where Mongolic has a bisyllable, apparently a derivative: 𐲃 *ku* †*ku* 'person' (genitive 𐲃𐲅𐲊 *ku.û.un*) = **kü-xün* id.,[32] 𐲋𐲍 *n.o* †*no* 'spouse' = **nö-kür* id., 'friend.' In the item 𐲋𐲎𐲏 *n.ai.ci* †*naiji* 'friend' = **nai-ji* id. both branches have a derivative; the root **nai* 'friendship' is preserved in Mongolic, and it is also attested in other derivatives like **nai-r* 'concord, festivity.'

Root structure. Khitan seems to have been characterized by the same type of basic phonotactic restrictions as the other "Altaic" languages, including both Jurchen and Proto-Mongolic. This means, among other things, that there were no syllable-initial or syllable-final consonant clusters. Most lexical items are bisyllabic ((C)V(C)CV(C)), but there are also conspicuously many monosyllables. Among the latter, those ending in a consonant ((C) VC) can have direct cognates in Mongolic, as in the case of 𐲈𐲌 *go.er* 'tent' > 'household' = **ger*.[33] Khitan has, however, also many monosyllabic vowel stems ((C)V), a type absent in Proto-Mongolic (except in pronouns). In the comparative context it seems that the monosyllabic vowel stems in Khitan are an archaic trait eliminated on the Mongolic side by way of secondary suffixation, as in the above-mentioned items †*ku* 'person,' †*no* 'spouse.'

Importantly, the method of internal reconstruction allows monosyllabic vowel stems to be established also for a very early stage of the Mongolic lineage, as in the collective nouns formed by **-d* (for countables) and **-s* (for liquifiables). Only one of these is so far identified in the Khitan Small Script database: 𐲋𐲔𐲕 *ci.i.is* †*cis* 'blood' > 'blood relative' = **ci-s-ü/n* 'blood,' presupposing the monosyllabic root **ci*. As may be seen, this item

[31] The only problem in the item for 'younger brother' is that the Romanization of the character is not verified by independent information. The genitive form confirms, however, that the underlying syllable ended in †*u*, suggesting that the Romanization is correct.

[32] Mongolic **küxün* has the variant **kümün* (as in Written Mongol and Oirat), suggesting that the original shape was **kü-pün* (with early dialectal nasalization of the medial consonant).

[33] The exact "reading" of the Khitan word is difficult to reconstruct. The comparative evidence would suggest the shape †*ger*, but the orthographical image is exceptional and might imply something more complex, especially as far as the vowel quality is concerned.

shows the collective suffix *-*s* also in Khitan, suggesting that this group of nouns was formed already in the common protolanguage of the Proto-Mongolic and Para-Mongolic lineages.[34]

5. Diachronic phonology

No systematic picture of Khitan phonology is yet possible to draw. We cannot, for instance, reconstruct the exact configuration of the systems of vowels and consonants. This is due to the nature of the script, which involves multiple cases of both under- and over-differentiation, as well as a considerable amount of orthographical variation. We also have to reckon with the possibility that Khitan changed in the course of time: the currently available corpus of Khitan Small Script texts covers a period of over a hundred years (c. 1041-1171),[35] while some of the orthographical conventions are likely to date from the time when the script was created (924-925). In general, work on the Khitan phonological system is closely connected with the reconstruction of Liao Chinese, a little known form of speech, transitional between local Late Middle Chinese and Early Mandarin.[36] On the other hand, Khitan phonology is likely to share many features with Jurchen on an areal basis.

In the following some general points of relevance in the diachronic context are discussed in comparison with the Proto-Mongolic lineage:

Vowel Rotation. As a working hypothesis it is reasonable to assume that Khitan originally had a vowel system similar to Pre-Proto-Mongolic. This system would have comprised 8 vowels, divided into 4 back (**a *o *u *i*) and 4 front (**e *ö *ü *i*), as well as 4 lower (**a *e *o *ö*) and 4 higher (**ï *i *u *ü*) vowels. A similar system, with minor modifications, may be

[34] On the Pre-Proto-Mongolic collectives in **-d* resp. **-s*, cf. Janhunen (1996: 210-215). On the Mongolic side, these nouns are always expanded by the element **-U/n*. It may be noted that the item for 'blood' is also the base for the Khitan term for 'filial piety,' see Kane (2004), Shimunek (2007: 71).

[35] These datings are based on the most recent summary of Khitan Small Script epitaphic documents by Wu Yingzhe (2012).

[36] On Liao Chinese, cf. Kane (2009: 227-264). Tentative outlines on Khitan phonology are given by Shimunek (2007: 38-52) and Takeuchi (2007).

reconstructed for Proto-Turkic, Proto-Tungusic, and even Proto-Koreanic. All over the region, the system has, however, undergone a process of vowel rotation, in which the original front vowels have been raised and velarized, while the original back vowels have been lowered and, in some languages, pharyngealized. A typical example of the effect of vowel rotation is the development of the pair **u* - **ü* = phonetically [u - y] to *o* - *u* = [o - u], as observed in languages such as Manchu, Dagur, and Korean.[37]

Vowel rotation in Northeast Asia is a relatively recent phenomenon, not yet observable with certainty in Proto-Mongolic. It is, therefore, important to note that it seems to have been present in Khitan. Although the vowel letters contained in the Romanizations should not be taken at face value, there are examples suggesting the developments **ü* > †*u* and **u* > †*o*, as in 杏 *uni* †*uni* 'ox' = **üni-xe/n* 'cow,' 夭 *on-* †*on-* 'to ride' = **unu-* id. The ultimate impact of vowel rotation is the reduction of the paradigm towards a simple five-vowel system (*a e i o u*).[38] We do not know whether Khitan had reached this stage, but there are indications of, at least, a neutralization between **ö* and **ü*, cf. 今杏 *s.uni* 'night' = **söni* id. On the other hand, there is also evidence of a more general neutralization of rounded vowels, cf. e.g. the numerals 3 刘化 *hu.ur-* †*hur-* = **gur-* vs, 4 令化 *t.ur-* †*dur-* = **dör-*.[39]

*Preservation of *p*. Another widespread areal feature in Northeast Asia concerns the original strong (voiceless/aspirated) labial stop (**p*), which has been variously spirantized (> *f*), velarized (> *x*), pharyngealized (> *h*), or lost (> Ø) in almost all languages and language families of

[37] For a general discussion of the areal and typological background of vowel rotation in the languages of Northeast Asia, cf. Janhunen (1981). This model of vowel rotation has recently been challenged by Ko Seongyeon (2011), who claims that the diachronic process took place in the opposite direction, that is, from an apertural and/or pharyngeal (RTR = "retracted tongue root") harmony to palatal harmony. Although the issue is of minor relevance to the number of oppositions in the paradigm, and the two types of vowel harmony (palatal and apertural) coexist in Northeast Asia, the arguments presented in favour of the revised model do not seem to be tenable.

[38] The system can be secondarily increased by the introduction of a new set of palatal vowels, as has happened in Korean, as well as in several varieties of Mongolic.

[39] In numerals we have, of course, to consider also the possibility of analogical levelling between consecutive items.

the region, including Turkic, Mongolic, Tungusic, and even Japonic.[40] While vowel rotation seems to have proceeded from east to west, the spirantization of the strong labial stop moved in the opposite direction, being most ancient in the west (Turkic) and most recent in the east (Tungusic and Japonic). There is only one language family of the "Altaic" sphere, Koreanic, in which this phenomenon is totally absent.[41]

Importantly, the strong labial stop is also preserved in Khitan. The same seems to have been the situation in early Jurchen, while spirantization (> *f*) took place in later Jurchen, as attested in Manchu. The status of Khitan in this respect was known already from the lexical data preserved in Chinese rendering, as in 頗 †*po* = 非 *po* 'time' = Pre-Proto-Mongolic **pon* (= Jurchen †*pon*) > Proto-Mongolic **xon* > Modern Mongolian *on* 'year.'[42] Today, we have a few more examples, including 北本 *p.ar* †*par*(*a*) 'people' = **para* > **xara/n* : (plural) **xara-d* > *ar*(*a*)*d* id., 北平久 *p.ul.uh* †*pulu*(*-h*) 'intercalary month' < 'more' = **püle-xü* 'more, extra,' as well as 北卖今 *p.úr.s* †*pur*(*e*)*s* 'descendants' = (plural) **püre-s* 'seeds.'

*Spirantization of *k*. Very much like the strong labial stop **p*, the corresponding velar stop **k* has also tended to undergo spirantization (> *x*) and pharyngealization (> *h*), more rarely loss (> Ø), in the languages of Northeast Asia. This tendency is typically older and/or more prominent in those languages that preserve, or have until recently preserved, **p* intact. Therefore, we find the development **k* > *h* as a relatively early phenomenon in languages like Korean and Jurchen-Manchu,[43] while in the

[40] The phenomenon is also attested in a few Uralic languages. Starting with Pelliot (1925), particular attention in comparative Altaic studies has been paid to the *h*-stage, though the implications of the issue have not always been correctly understood.

[41] Another language family in which the phenomenon is not attested is Amuric, today represented by the single language (isolate) of Ghilyak (Nivkh), but earlier distributed more widely in Manchuria, including, quite possibly, the northern neighbourhood of Korean.

[42] On this item, cf. Kane (2009: 68, 122-123). The Jurchen word is, of course, a borrowing from some early Mongolic form of speech. It may be noted that the Khitan "reading" is †*po*, while Mongolic and Jurchen suggest the presence of a final nasal, i.e. **pon*. There are also other cases in which Khitan and the Proto-Mongolic lineage differ in the use of the final nasal **n* in nominals. The reasons underlying these differences are not known.

[43] The spirantization of **k* in Jurchen-Manchu has been the object of some dispute. However, the development seems to have been regular in initial position, as maintained

Proto-Mongolic lineage (as well as in Turkic) it is much later. Depending on the language, the development may have contextual restrictions; it is particularly common that it is present only before original back vowels.

In Khitan, we have, in fact, examples of preserved **k* before original front vowels, as in 𐲁 *ku* †*ku* 'person' (genitive 𐲁𐲰𐲭 *ku.û.un*) = **kü-xün* id. The variation in 𐲢𐲴 *k.iú* ~ 𐲵𐲴 *x.iú* †*keu* 'younger sister' = **keü* 'child' might suggest that there was a tendency to spirantize **k* also in this position, another possible example being 𐲵𐲴𐲇 *x.ui.ri* 'to arrive,' which may or may not correspond to Proto-Mongolic **kür-* id.[44] Before original back vowels, spirantization is confirmed by 𐲀𐲶 *heu.úr* †*haur* = **kabur* 'spring.' Again, caution should be exercised with regard to the letters used in the Romanization (<k x h>), the exact phonetic values and phonemic relationships of which are not known. It may be noted that, at least at the level of Romanization, a spirantized velar in Khitan can also correspond to the Mongolic weak velar stop **g*, as in 3 𐲉𐲇 *hu.ur-* †*hur-* = **gur-*.

Vowel contraction. The Proto-Mongolic lineage had a medial velo-laryngeal spirant **x* [h], which is represented as a segment in Written Mongol, but which has been lost in all extant Mongolic languages, and even in Middle Mongol, with vowel contraction (long vowels and diphthongs) as a result. The segment itself was of a heterogeneous origin and may in some cases even have been "original."[45] The issue is complicated by the possibility that there may also have existed "original" diphthongs (like **ai* **au*), though the comparative material does not allow them to be distinguished reliably from sequences with a medial **x* (like **axi* **axu*).

Proto-Mongolic **x* is, in any case, normally represented as zero in what may probably be understood as diphthongs in Khitan, as in 100 𐲈 *jau* †*jau* = **jaxu/n*. However, the segmental counterpart of Proto-Mongolic **b* also seems to be lost in Khitan when followed by a rounded vowel, as in 5 𐲆 *tau* †*tau* = **tabu/n*, 𐲀𐲶 *heu.úr* †*haur* = **kabur* 'spring.' The fact that there is no medial consonant in these cases is confirmed by the use of

---

by Näher (1999); for a different explanation, cf. Vovin (1997).

[44] This correspondence would be valid if the Khitan item is to be read †*hur-*, but the peculiar orthographical image may actually point to a different shape. In any case, there must have been a difference with regard to the numeral 3 †*hur-*.

[45] On **x* in Mongolic, cf. Janhunen (1999). In earlier research, this segment was often confused with **g*, with which it is graphically indinguishable in Written Mongol.

the characters 𘰾 *jau* and 𘰿 *tau* for the Chinese term 𘰾𘰿 *jau.tau* †*jautau* 'punitive commissioner' = 招討 **jawtaw* (*zhaotao*). The use of the character 𘰿 *tau* in the word 𘰿𘱁𘱂 *tau.li.a* 'hare' = **taula-i* id. is also significant, though the phonological background of this "rebus" remains unclear.[46]

The fate of **x* in monophthongoid contexts in Khitan is less clear, which is why comparisons like 𘰽 *ún* †*un* = ?**e-xü/n-* are uncertain. A possible example would be offered by 𘰼 *qa* 'emperor,' but this item is ambiguous, since it could correspond to either **ka.n* 'prince' or **kaxa.n* 'emperor' on the Mongolic side. The genitive 𘰼𘰼𘱇 *qa.ha.án* 'of the emperor' is also enigmatic since it seems to contain a medial **g*, as in 𘰺𘰼 *na.ha* †*naha* 'maternal uncle' = **naga-cu* id. It is possible that the actual Khitan readings should be †*kaga* : †*kaga-n*, in which case the Khitan word would correspond to Turkic **kaga.n*, rather than to Mongolic **kaxa.n*. Altogether, this is a word that seems to involve an unknown network of borrowings.

Vowel elision. The loss of vowels in non-initial syllables is a trivial process observed in many languages. The phenomenon may concern all short (single) vowels, or only those in certain positions, or those with certain qualities. In Northeast Asia, an early vowel elision may be established for several language families, including Turkic and Korean(ic), while other families, including Mongolic and Tungusic, show a less systematic, or at least chronologically later, tendency to lose vowels.[47] A special feature of Jurchen-Manchu (and certain forms of Mongolic) is the loss of vowels in open second syllables (the so-called "Mittelsilbenschwund").

---

[46] The issue depends on what part of the Proto-Mongolic numeral Khitan †*tau* corresponds to: only to the root **tab*, or to the entire bisyllabic sequence **tabu*. The word **taulai*, on the other hand, is one of the items containing an "original" diphthong with no medial consonant in Written Mongol, a situation further complicated by the fact that the Turkic counterpart **tabïsh-gan* has a medial **b*. However this may be, the connection between 'five' and 'hare' was among the strongest pieces of evidence in favour of the Mongolic identification of Khitan during the early phases of the decipherment of the Khitan Small Script, cf. e.g. Kara (1975: 165-166). On the diachrony of the Turko-Mongolic item for 'hare,' cf., most recently, Rybatzki (2010: 188-189).

[47] The diachronic fact of vowel elision in a language can be easily established and demonstrated by external comparisons with other languages with preserved vowels. In Turkic and Korean, for instance, the phenomenon is confirmed by comparisons such as Turkic **er* 'man' = Mongolic **ere*, Korean *kom* 'bear' = Japanese *kuma*. Another language that has lost vowels is Ghilyak, as may be seen from examples such as Ghilyak *camng* 'shaman' = Manchu *saman*. Note that all these examples involve areal contacts (loanwords), rather than cases of genetic relationship.

In Khitan, the nature of the script makes it often difficult to verify the presence or absence of vowels in all positions, including even word-initially. Many Khitan characters, irrespective of what Romanizational shape is used for them, refer basically to a consonantal core, which may be preceded (VC) and/or followed (CV) by a vowel. Items such as 𘥈𘤏 *b.as* 'again, also' = **basa* id. and 𘤟𘤽 *p.ar* 'people' = **para-* id. may, therefore, well have contained a final vowel in Khitan. On the other hand, the presence of a vowel letter in the Romanization does not necessarily mean that there was a segmental vowel in the Khitan word. We cannot, therefore, be certain that items such as, for instance, 𘥣𘠽 *i.ri* 'name' = **nere* id. (with a secondary initial nasal) and 𘥔𘠽 *m.ri* 'horse' = **mori/n* id. ended in a vowel.

Even so, the circumstance that final vowels are retained in Jurchen-Manchu suggests that Khitan most probably also retained them. This is, incidentally, also suggested by the fact that certain suffixes ending in a vowel (-CV) have positionally alternating variants, although it is not necessarily a question of regular vowel harmony.[48] It is also important to note that the counterparts of Mongolic long vowel elements (diphthongs and contracted vowels) are in Khitan rendered by using additional vowel letters, as in 𘤪𘤱𘥋 *t.qo.a* 'chicken' = **takï-xa* id., 𘥑𘤿𘤁 *mu.ho.o* 'snake' = **moga-i* id., and also 𘥖𘥗𘥋 *tau.li.a* 'hare' = **taula-i* id. For the time being we cannot reconstruct the Khitan shapes of such items with any certainty, but the orthography suggests that they may have involved complex vowel elements (diphthongs or long vowels) also in Khitan.

*

It has to be stressed that only a small selection of diagnostic lexical and phonological properties of Khitan has been discussed above. Much more remains to be said, and even more remains to be discovered. What is, however, certain is that the material is consistent with the status of Khitan as a Para-Mongolic language. The distance between Khitan and Proto-Mongolic is clearly large enough to have made immediate communication between the two speech communities impossible. At the same, the

[48] An example is offered by the dative case ending, Romanized as *-de* ~ *-do* ~ *-du*, cf. Kane (2009: 136-138). The nature of "vowel harmony" in Khitan is still an open issue.

similarities are conspicuous enough to have been evident even for the naïve native speaker. The real situation may have been modified by the presence of intermediate idioms of which we have no direct information.

As far as the position of Khitan on the archaic—innovative scale is concerned, the situation is ambiguous. Even so, at least phonologically, Khitan may be characterized as an innovative language, which in most respects had evolved more rapidly than the lineage leading to Proto-Mongolic. Many areal innovations reached Khitan earlier than the Proto-Mongolic lineage. The reasons for the different speed of evolution must be searched in the geographical, demographical and political situation. In the centuries preceding the Liao period, Khitan had become the language of a rapidly growing and highly mobile population, which, moreover, contacted on a wide scale with a variety of other speech communities. For the Proto-Mongolic lineage, such a period of intensive growth and contacts was yet to come.

References

Alonso de la Fuente, José (2011). *Tense, Voice and* Aktionsart *in Tungusic: Another Case of "Analysis to Synthesis"?* Tunguso-Sibirica 32. Wiesbaden: Harrassowitz Verlag.

Arapov, M. V. (1982). Leksika i morfologiia tekstov malogo kidan'skogo pis'ma. In: *Zabytye sistemy pis'ma: Materialy po deshifrovke*. Moskva: Izdatel'stvo "Nauka." 211-239.

Chingeltei 清格爾泰 [Qinggeertai] (1997). Qidanyu shuci ji Qidan Xiaozi pindufa 契丹語數詞及契丹小字拼讀法. *Altai Hakpo* **7**. 143-152.

Chingeltei 清格爾泰, Liu Fengzhu 劉鳳翥, Chen Naixiong 陳乃雄, Yu Baolin 于寶林, and Xing Fuli 邢復禮 (1985). *Qidan Xiaozi Yanjiu* 契丹小字研 究. Beijing 北京: Zhongguo Shehui Kexue Chubanshe 中國社會科學出版社.

Doerfer, Gerhard (1985). *Mongolo-Tungusica.* Tungusica 3. Wiesbaden: Otto Harrassowitz.

Doerfer, Gerhard (1992). Mongolica im Alttürkischen. In *Bruno Lewin zu Ehren* 3. Bochum. 30-56.

Doerfer, Gerhard (1993). The Older Mongolian Layer in Ancient Turkic. *Türk Dilleri Araştırmaları* 1993:3. 79-86.

Golovachev, V. C., A. L. Ivliev, A. M. Pevnov, and P. O. Rykin (2011).

Tyrskie stely XV veka: Perevod, kommentariï, issledovanie kitaiskix, mongol'skogo i chzhurchzhen'skogo tekstov. Sankt-Peterburg: "Nauka."

Grube, Wilhelm (1896). *Die Sprache und Schrift der Jučen*. Leipzig: Kommissions-Verlag von O. Harrassowitz [Reprints: Peking 1936, Tientsin 1941].

Janhunen, Juha (1981). Korean Vowel System in North Asian Perspective. *Han-geul* **172**. 129-146.

Janhunen, Juha (1993). The Teens in Jurchen and Manchu Revisited. In *Festschrift für Raija Bartens*. Ulla-Maija Kulonen (ed.) Mémoires de la Société Finno-Ougrienne 215. Helsinki. 169-184.

Janhunen, Juha (1994). On the Formation of Sinitic Scripts in Mediaeval Northern China. *Journal de la Société Finno-Ougrienne* **85**. 107-124.

Janhunen, Juha (1995). Kittanjin ha nanigo wo hanashite ita ka 契丹人はなに語をはなしていたか. *Minpaku Tsuushin* 民博通信 **68**. 82-85.

Janhunen, Juha (1996). Prolegomena to a Comparative Analysis of Mongolic and Tungusic. In *Proceedings of the 38th Permanent International Altaistic Conference* (PIAC) (Kawasaki, Japan: August 7-12, 1995). Giovanni Stary (ed.). Wiesbaden: Harrassowitz Verlag. 209-218.

Janhunen, Juha (1999). Laryngeals and Pseudolaryngeals in Mongolic: Problems of Phonological Interpretation. *Central Asiatic Journal* **43**. 115-131.

Janhunen, Juha (2003). Para-Mongolic. In *The Mongolic Languages*, Juha Janhunen (ed.), Routledge Language Family Series 5. Routledge: London & New York. 391-402.

Janhunen, Juha (2012). The Expansion of Tungusic as an Ethnic and Linguistic Process. In *Recent Advances in Tungusic Studies*, Andrej L. Malchukov & Lindsay J. Whaley (eds.) Turcology 89. Wiesbaden: Harrassowitz Verlag. 5-16.

Jin Qicong 金啓孮 (1984). *Nüzhen wen cidian* 女真文辞典. Beijing 北京: Wenwu Chubanshe 文物出版社.

Kane, Daniel (1989). *The Sino-Jurchen Vocabulary of the Bureau of Interpreters*. Indiana University Uralic and Altaic Series 153. Bloomington.

Kane, Daniel (2004). A Note on *Čisdeben. *Central Asiatic Journal* **48**. 223-225.

Kane, Daniel (2006). Khitan and Jurchen. In Alessandra Pozzi, Juha

Janhunen, and Michael Weiers (eds.) *Tumen Jalafun Jecen Akū: Manchu Studies in Honour of Giovanni Stary*. Tunguso-Sibirica 20. Wiesbaden: Harrassowitz Verlag. 121-132.

Kane, Daniel (2008). A Transcription System for Texts in the Qidan Small Script. Handout presented at International Conference on Qidan, Tangut, Jurchen and Mongolian Historical Documents 遼夏金元歷史文獻國際研討會 *Liao Xia Jin Yuan Lishi Wenxian Guoji Yantaohui* (Beijing, 2008.11.3-5).

Kane, Daniel (2009). *The Kitan Language and Script*. Handbook of Oriental Studies VIII, 19. Brill: Leiden & Boston.

Kara, György (1975). Apropos de l'inscription khitane de 1150. *Annales Universitatis Scientiarum Budapestinensis*. Sectio Linguistica **6**. 163-167.

Kiyose, Gisaburo N. (1977). *A Study of the Jurchen Language and Script: Reconstruction and Decipherment*. Kyoto: Hôritsubunka-sha.

Ko Seongyen (2011). Vowel Contrast and Vowel Harmony Shift in the Mongolic languages. *Language Research* **47:1**. 23-43.

Kwanten, Luc (1984). The Phonological Hypothesis of the Hsi Hsia (Tangut) Language. *T'oung Pao* **70**. 159-184.

Kwanten, Luc (1988). The Structure of the Tangut (Hsi Hsia) Characters. *Journal of Asian and African Studies* **36**. 69-105.

Ligeti, Louis (1970). Le tabghatch, un dialecte de la language sien-pi. In Louis Ligeti (ed.) *Mongolian Studies*. Bibliotheca Orientalis Hungarica 14. Budapest. 265-308.

Menges, Karl H. (1968). *Tungusen und Ljao*. Abhandlungen für die Kunde des Morgenlandes 38:1. Wiesbaden: Franz Steiner GmbH.

Näher, Carsten (1999). Der urtungusische stimmlose velare Plosiv im Mandschu. *Journal de la Société Finno-Ougrienne* **88**. 113-130.

Pelliot, Paul (1925). Les mots à *h* initiale, aujourd'hui amuie, dans le mongol des XIIIe et XIVe siècles. *Journal Asiatique* **206**. 193-263.

Pevnov, A. M. (2004). *Chtenie chzhurchzhen'skix pis'men*. Sankt-Peterburg: Rossiiskaia Akademiia Nauk, Institut lingvisticheskix issledovanii & "Nauka."

Róna-Tas, András (2004). A Khitan Word for Marmot. *Acta Orientalia Academiae Scientiarum Hungaricae* **57**. 27-29.

Rybatzki, Volker (2003). Middle Mongol. In Juha Janhunen (ed.) *The Mongolic Languages*, Routledge Language Family Series 5. Routledge:

London & New York. 57-82.

Rybatzki, Volker (2011). Classification of Old Turkic loanwords in Mongolic. In Mehmet Ölmez, Erhan Aydın, Peter Zieme, and Mustafa S. Kaçalin (eds.) *From Ötüken to Istanbul: 1290 Years of Turkish*. Istanbul Büyükşehir Belediyesi. 185-202.

Shimunek, Andrew A. (2007). *Towards a Reconstruction of the Kitan Language, with Notes on Northern Late Middle Chinese Phonology*. MA Thesis, Indiana University, Department of Linguistics & Department of Central Eurasian Studies.

Starikov, V. S., M. Arapov, A. Karapet'ianc, Z. Malinovskaia, and M. Probst (1970). *Materialy po deshifrovke kidan'skogo pis'ma* 1-2. Moskva: Akademiia Nauk SSSR.

Sun Bojun 孫伯君 & Nie Hongyin 聶鴻音 (2008). *Qidanyu Yanjiu* 契丹語研究, Zhongguo Shehui Kexueyuan Wenku Wenxue Yuyan Yanjiu Xilie 中國科學院文庫文學語言研究系列. Beijing 北京.

Takeuchi Yasunori 武內康則 (2007). *Kittan Shouji de Hyouki Sareta Kanjion kara Mita Kittango On'in Taikei no Kenkyuu* 契丹小字で表記された漢字音から見た契丹語音韻体系の研究. MA Thesis. Kyoto University: Kyoto Daigaku Daigakuin Bungaku Kenkyuuka 京都大學大學院文學研究科.

Talpe, Lode (2010). Some Qidan Words in Chinese Poems. *Central Asiatic Journal* **54**. 79-91.

Toyoda Gorou [Gorō] (1964). An Analysis of the Major Ch'i-tan Characters. *Memoirs of the Research Department of the Toyo Bunko* **23**. 119-135.

Toyoda Gorou [Gorō] 豐田五朗 [Fengtian Wulang] (1998). Qidan Xiaozi dui siji de chenghu 契丹小字对四季的称呼. *Minzu Yuwen* 民族語文 1998:1. 78-81.

Vovin, Alexander (1997). Voiceless Velars in Manchu. *Journal de la Société Finno-Ougrienne* **87**. 263-280.

Vovin, Alexander (2003). Once Again on Khitan Words in Chinese-Khitan Mixed Verses. *Acta Orientalia Academiae Scientiarum Hungaricae* **56**. 237-244.

Vovin, Alexander (2011). A Modest Proposal on the Decipherment of the Khitan-Chinese Bilingual Text of 1134 (the Langjun Inscription). In Michael Knüppel & Aloïs van Tongerloo (eds.). *Life and Afterlife & Apocalyptic Concepts in the Altaic World*, Proceedings of the 43rd Annual Meeting of the Permanent International Altaistic Conference (PIAC), Tunguso-Sibirica 31. Wiesbaden: Harrassowitz Verlag. 123-

130.

Wittfogel, Karl A. & Fêng Chia-shêng (1949). *History of Chinese Society: Liao* (*907-1125*). Transactions of the American Philosophical Society. New Series 36. Philadelphia.

Wu Yingzhe 吳英喆 (2005). Qidan Xiaozi «xing» de yufa fanchou chutan 契丹小字 "性" 的語法範疇初探. *Nei Menggu Daxue Xuebao* (*Zhexue Shehui Kexueban*) 內蒙古大學學報 (哲學社會科學版) 2005:3. 25-28.

Wu Yingzhe (2007). *Qidanyu Jingci Yufa Fanchou Yanjiu* 契丹語靜詞語法範疇研究 [Research on the grammatical categories of the noun in the Khitan language]. Huhe-haote 呼和浩特: Nei Menggu Daxue Chubanshe 內蒙古大学出版社.

Wu Yingzhe (2009). A Brief Discussion on the Vowel Attachment in the Khitan Small Script. *Voprosy filologii, Seriia "Uralo-altaiskie issledovaniia"* **1**. 26-30.

Wu Yingzhe (2011). Deciphering Some Demonstrative Pronouns in Khitan Small Script. *Altai Hakpo* **21**. 69-77.

Wu Yingzhe (2012). Organization of the Khitan materials kept in Inner Mongolia, Unpublished Powerpoint presentation (Tokyo University of Foreign Studies). 1-67.

Wu Yingzhe & Juha Janhunen (2010). *New Materials on the Khitan Small Script: A Critical Edition of Xiao Dilu and Yelü Xiangwen*. Corpus Scriptorum Chitanorum 1, The Languages of Asia Series. Folkestone, Kent: Global Oriental.

Juha A. Janhunen
Institute for Asian and African Studies
University of Helsinki, Finland
<asiemajeure@yahoo.com>

[Received 10 May 2012;
accepted 10 August 2012]