*Proceedings*

*of the SCRIPTA 2011*

# Writings and Cognition



Date: October 7~9, 2011

Venue: Sinyang Humanities Hall,
Seoul National University, KOREA



훈 민 정 음 학 회

The Hunmin jeongeum Society

Copyright © The Hunmin jeongeum Society

Published in 2011



훈 민 정 음 학 회
The Hunmin jeongeum Society
c/o Dept. of Linguistics, Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul, 151-745, Korea
Tel. +82-2-880-6167, 6168 Fax. +82-2-882-2451
e-mail: scripta08@gmail.com

*Proceedings of the SCRIPTA 2011*

Date: October 7~9, 2011
Venue: Sinyang Humanities Hall, Seoul National University, KOREA

# Contents


Program ····· iii

Welcome address ····· *Doo Hyun PAEK* ····· v

Congratulatory Address ····· *LEE Ki-nam* ····· vii

Tokens in China, Europe and Africa

····· *Denise SCHMANDT-BESSERAT* ····· 1

Discussion ····· *Seung Il KANG* ····· 15

Khitan: Understanding the language behind the scripts

····· *Juha A. JANHUNEN* ····· 19

Discussion ····· *LEE Seong Gyu* ····· 49

Categorization of meaning in the Chinese ideographic characters

····· *Jungsam YUM* ····· 51

Discussion ····· *Olivia MILBURN* ····· 69

Hieroglyphic writing in ancient Egypt and its cognitive involvement

····· *Pascal VERNUS* ····· 73

Discussion ····· *Sung Hwan YOO* ····· 111

A proposal for a Hunminjeongeum-based writing system for Aymara

····· *SeungJae LEE, Juwon KIM, Sangchul PARK, Hui JIN, Chang Min KIM* ····· 115

Discussion ····· *Seong Kyu KIM* ····· 143

*Tongguk chŏngun* and the phonological system of Middle Korean

····· *Rei FUKUI* ····· 147

Discussion ····· *CHA Ikjong* ····· 163

The rebirth of Hangeul

····· *KIM Young-wook* ····· 167

Discussion ····· *KIM Sang-Tae* ····· 179

# Program

## October 7, Friday

18:00-20:00 Welcoming reception (Hoam Faculty House)

## October 8, Saturday

09:00-09:30 Registration & opening ceremony (Sinyang Humanities Hall)

Moderator: *PARK See-Gyoon* (Kunsan Nat'l Univ.)

Welcome address: *PAEK Doo Hyun* (President, the Hunmin jeongeum Society)

Congratulatory address: *LEE Ki-nam* (Chairman of the board, the Hunmin jeongeum Society)

**Session 1**

Moderator: *PARK See-Gyoon* (Kunsan Nat'l Univ.)

09:30-10:15 Tokens in China, Europe and Africa: *Denise SCHMANDT-BESSERAT* (University of Texas at Austin)
[Discussant: *KANG Seung Il* (HannamUniv.)]

10:15-10:30 Coffee break

Moderator: *MOK Jungsoo* (Univ. of Seoul)

10:30-11:15 Khitan: Understanding the language behind the scripts: *Juha A. JANHUNEN* (University of Helsinki)
[Discussant: *LEE Seong Gyu* (Dankook Univ.)]

11:15-12:00 Categorization of meaning in the Chinese ideographic characters: *YUM Jungsam* (Seoul Nat'l Univ.)
[Discussant: *Olivia MILBURN* (Seoul Nat'l Univ.)]

12:00-14:00 Lunch (Jahayeon)

**Session 2**

Moderator: *Marc DUVAL* (Seoul Nat'l Univ.)

14:00-14:45 Hieroglyphic writing in ancient Egypt and its cognitive involvement: *Pascal VERNUS* (École Pratique des Hautes Études)
[Discussant: *YOO Sung Hwan* (Brown Univ.)]

14:45-15:30 A proposal for a Hunminjeongeum-based writing system for Aymara: *LEE SeungJae, KIM Juwon, PARK Sangchul, JIN Hui, KIM Chang Min* (Seoul Nat'l Univ.)
[Discussant: *KIM Seong Kyu* (Seoul Nat'l Univ.)]

15:30-15:45 Coffee Break

Moderator: *JUN Youngchul* (Seoul Nat'l Univ.)

15:45-16:30 *Tongguk chŏngun* and the phonological system of Middle Korean: *Rei FUKUI* (University of Tokyo)
[Discussant: *CHA Ikjong* (Seoul Nat'l Univ.)]

16:30-17:15 The rebirth of Hangeul: *KIM Young-wook* (Univ. of Seoul)
[Discussant: *KIM Sang-Tae* (Cheongju University)]

18:00-20:00 Reception (Sodam Maru)

**October 9, Sunday:** Academic tour of National Museum of Korea
(Ancient writings and thereafter)

# Welcome address

Doo Hyun PAEK
(President, The Hunmin jeongeum Society)

Ladies and gentlemen, and eminent scholars from all over the world, thank you very much for attending this conference. I would like to express my sincere gratitude for your participation as both presenters and discussants. I am especially honored to have many leading international scholars in the field of linguistics at this conference to discuss various aspects of the alphabetic systems of the world.

It is my sincere pleasure to host the Third International Conference with the same theme of "Writings and Civilizations" as in 2010. Today, the 9th of October, is the very day for commemorating the creation of Hangeul. It is particularly meaningful for the Hunmin jeongeum Society to hold this international conference on this day.

To briefly report on the history of the Hunmin jeongeum Society, it was founded in 2007 to investigate various writing systems of the world, including Hangeul. The first international conference hosted by this Society was held on October 9th of 2008. *Scripta* vol. 2 was issued in September 2010 as a result of that first international conference. It is our hope that this publication will develop to become one of the major journals

dealing with the writing systems of the world. I believe that today's third international conference will produce even more significant achievements.

In addition, this Society has made strenuous efforts to preserve the endangered languages for the sake of language diversity. This Society has drawn worldwide attention by having Hangeul adopted as a way of recording Tsia-tsia, a language spoken by a minority tribe living on Buton island in Indonesia. In 2011 this Society is working on how to transcribe the native Aymara language of Bolivia employing the Hangeul alphabet.

Today I am grateful to all of you for your participation. I am particularly grateful for those scholars who have come from overseas as well as from this country to present your papers. I am also thankful to those scholars who have willingly agreed to serve as discussants.

I am confident that the discussions in this conference will prepare the way for further in-depth investigations of the alphabetic systems of the world, hence obtaining a more insightful understanding of writing systems. The presentations and discussions throughout this conference will certainly lead to successful and meaningful results.

Thank you very much.

# Congratulatory Address

LEE Ki-nam
Chairman of the board, the Hunmin jeongeum Society

Distinguished Ladies and Gentlemen,

I want to say congratulations on having the third international symposium.

I stand up before this distinguished guests today to speak briefly about the accomplishments to date and purposes of establishment of Hunmin jeongeum Society, as my term as chairperson of the organization expires today.

Hunmin jeongeum Society, an incorporated association was founded in 2007. Its founding purposes include the first task of publishing an international journal on script and secondly sharing alphabets using Jeongeum with people without their own written languages.

My role is to financially support the society so it will develop continuously. And it has been through Wonam Foundation, an incorporated foundation which I organized and named after pseudonym of my late father.

In the meantime, SCRIPTA vol. 3 was issued and international academic conference that was hosted every two years is now being held annually.

Then, i) script was shared with CiaCia people, who had no written language, which resides in Buton island, Indonesia, ii) script for Aymara

people in Bolivia, a South American country was made, and iii) preparation is on the way to provide written alphabets to a minority group of Zimbabwe in SE Africa.

Today, scholars of Hunminjeongum Society Indonesia are joining us now. Looking back on 9th of last month, CiaCia people declared the day as 'Baubau Hangul Day' in commemoration of having their own script.

Hunmin jeongeum Society, using Hunmin jeongeum, creative invention of King Sejong the Great is dedicating its efforts for humankind as well as for study. I would like to ask your active participation and encouragement for the society's further development.

Thank you very much and God bless you.

# Tokens in China, Europe and Africa

## – The significance

Denise SCHMANDT-BESSERAT
The University of Texas at Austin, USA

This paper deals with the fact that tokens, familiar in the prehistoric Near East,[1] are also recovered in excavations in China, Europe and Africa and most recently in Mesoamerica.[2] Tokens were small clay artifacts in geometric shapes that were used as counters to keep track of goods before writing. I list below a selection of sites that yield tokens outside the Near

[1] Denise Schmandt-Besserat, *Before Writing*, University of Texas Press, Austin 1992; "Tokens and Writing: The Cognitive Development," *Scripta*, Vol. 1, 2009, P.145-154; "The Token System of the Ancient Near East: its role in Counting, Writing, the Economy and Cognition," in Iain Morley and Colin Renfrew, *The Archaeology of Measurement*, Cambridge University Press, 2010, P. 27-34.

[2] Linda R. Manzanilla, "Corporate Life in Apartment and Barrio Compounds at Teotihuacan, Central Mesxico," In Linda R. Manzanilla and Claude Chapdelaine, *Domestic Life in Prehispanic Capitals*, Memoirs of the Museum of Anthropology, University of Michigan, Ann Arbor 2009, P. 30, Fig. 2.8. (21-42).

East and consider what can be learned from the new data on the universality of the art of counting with geometric counters.

## China

Tokens were recovered in recent Chinese excavations in the present-day Anhui province of the People's Republic. The artifacts belong to the Neolithic Shuangdun Culture that flourished about 5500 BC.[3] The types most frequently used were cones, spheres and disks.[4] Further Neolithic Chinese sites yielding tokens are Qianshan Xuejiagang, also in the Anhui province, Lintong Jiangzhai in the Shanxi province, and Tianmen Dengjiawan in the Hubei province.[5]

## Europe

In Eastern Europe, assemblages of Neolithic tokens from Slovenia have been the subject of a careful study by Michael Budja.[6] Further west, tokens dated to early fourth millennium BC in the shapes of cones and spheres were identified in Italy at Capo Alfiere, Calabria.[7] Still further evidence

[3] Kan Xu Hang, Zhou Qun and Xu Da Li, eds., *Bengbu Shuangdun – A Report on the Neolithic Site*, Vols. 1-2, Anhui Provincial Institute of Antiquity and Archaeology, Bengbu Museum, Science Press, Beijing 2008.

[4] Kan Xu Hang, Zhou Qun and Xu Da Li, eds., *Bengbu Shuangdun – A Report on the Neolithic Site*, Vol. 2, Anhui Provincial Institute of Antiquity and Archaeology, Bengbu Museum, Science Press, Beijing 2008 P. 130: 1-12; 133: 1-4 and 132: 1-12.

[5] Yan Zhi and Park Zaifu, "Ancient Counting tools from Excavations – From Shang Zhai and other Cultures" in Wang Yuxin, et allii, eds., *Beijing and Chinese Civilization*, Social Science Archive Publisher, Beijing 2006. I am grateful to, Professor Huang Yaping, Ocean University of China, for this reference, and to Professor Zhang He, for the translation.

[6] Michael Budja, "Clay Tokens – Accounting before Writing in Eurasia," *Documenta Praehistorica* vol. XXV, 1998, P. 219-235.

[7] Jon Morter, "Four pieces of clay: 'tokens' from Capo Alfiere, Calabria." *Journal of Mediterranean Archaeology*, Vol. 7.1, 1994, p. 115-123.

from Vivara, Italy, in the first half of the second millennium is especially interesting because of a particular way to create subtypes by dividing disks into halves and quarters.

## Africa

Tokens are present in Egypt, but rare.[8] As was customary for the early Egyptian excavation reports, only the stone specimens were published at Abydos, namely 49 spheres.[9] Further south, in Khartoum, in the Sudan, A. J. Arkell reported two types of Neolithic tokens: spheres and disks with a number of subtypes bearing markings.[10] Ten of the 26 spheres reported display deep grooves. Disks feature incised lines on their face or around the edge. Jebel Moya, also in the Sudan,[11] yielded small cones and disks with various markings, as well as disks, parabollae and long shapes cut from potsherds dated from the Meroitic period at the end of the 1st millennium BC.[12]

## Mesoamerica

The Teotihuacan collection consists of small cones and cylinders modeled by hand,[13] and 550 disks made from potsherds or mica. The disks were

[8] Denise Schmandt-Besserat, "An Early Recording System in Egypt and the Ancient Near East," in Denise Schmandt-Besserat, ed., *Immortal Egypt*. Undena Publications, Malibu, California 1978, p. 5-12.

[9] Walter B. Emery, *Great Tombs of the First Dynasty*, Vol. II, Egypt Exploration Society, Oxford University Press, 1954, P. 56-59.

[10] A. J. Arkell, *Early Khartoum*, Oxford University Press, 1949, P. 79 ff.

[11] Frank Addison, *Jebel Moya*, vol. 1, Oxford University Press 1949, p. 214, 227, 241, 242 and Vol. 2, Plate CV: 11-13; CXIV: 6-11.

[12] M. Pitts, "Jebel Moya Revisited: a settlement of the 5th Millennium BC in the Sudan," *Antiquity*.ac.uk/Ant/065/0262/Ant0650262.pdf (January 9, 2011).

[13] Linda R. Manzanilla, "Corporate Life in Apartment and Barrio Compounds at Teotihuacan, Central Mesxico," In Linda R. Manzanilla and Claude Chapdelaine,

either whole, or cut into halves, quarters or thirds. Further substantial tokens assemblages were excavated at Teopancazco, where 12 % of the 530 disks were cut in halves, and at Xalasco, where 42% of the 303 disks were cut into quarters. The date of these tokens is 1-500 AD when Teotihuacan was a vast multiethnic metropolis in Central Mexico.

## The Three Token Groups

The presence of tokens in Shuangdun, western China, is not so surprising because, as shown in Jeitun[14] and Anau,[15] Central Asia produced in the 7th Milennium BC sizeable assemblages of tokens including cones, spheres, discs, and biconoids. The Shuangdun tokens can therefore be considered as the extension of the Central Asian prehistoric accounting system, which itself was indebted to Mesopotamia or Persia. If, as is the present consensus, agriculture was brought to Europe from the Near East, it is plausible to consider that the practice of counting with tokens traveled from East to West as part and parcel of the Neolithic tool kit.[16] Finally, it is more difficult, but still conceivable to argue, that the idea of clay counters was passed on from the Levant to Egypt and from there to Khartoum in the Sudan. In this perspective, the ancient Near East, including Syria, Anatolia, Mesopotamia and Persia, may be considered as the core region that initiated a system of counting and accounting with tokens, while China

---

*Domestic Life in Prehispanic Capitals*, Memoirs of the Museum of Anthropology, University of Michigan, Ann Arbor 2009, P. 30, Fig. 2.8. (21-42).

14 V.M. Masson and V. I. Sarianidi, *Central Asia*, Praeger Publishers, New York, 1972, P. 35.

15 Raphael Pumpelly, *Explorations in Turkestan, Prehistoric Civilizations of Anau*, Carnegie Institution, Washington, D.C., 1908, P. 167-168, figs. 395 and 396; Denise Schmandt-Besserat, *Before Writing*, Vol. 2, The University of Texas Press, Austin, Texas 1992, p. 28-31.

16 Bleda S. During, *The Prehistory of Asia Minor*, Cambridge University Press, Cambridge 2010.

and Europe were peripheral, with Egypt and Central Africa more tenuously so. Of course, neither the much later Meroitic tokens of Jebel Moya nor those of second millennium Vivara or of distant Teotihuacan can in any way be related to the core region or its periphery. They have, therefore, to be considered as a third group that I will refer to as “distant.”

## Similarities

The Chinese, European, and African tokens from Khartoum share material, forms and size with their Near Eastern prototypes. Most importantly, they were based on the same symbolism, and served a similar economic function.

It is not surprising that clay was the material chosen the world over to manufacture counters because, thanks to its remarkable quality of plasticity, it can easily be modeled, with the bare hand, in an infinite number of discreet shapes that are easy to recognize, identify, remember, and replicate. Cones, spheres, and disks are the most ubiquitous shapes showing a shared predilection for geometric shapes. In China and Europe the tokens are plain, i.e. mainly devoid of markings, which make them unequivocally similar to the Neolithic and Chalcolithic Near Eastern assemblages of 7500 -3500 BC, with the only difference that several forms such as cylinders, tetrahedrons and ovoids seem absent in the periphery.

From the origin, the Near Eastern tokens served to keep track of amounts of goods in the early agricultural communities. For instance, the earliest examples of 7500 BC were recovered in level III of the site of Mureybet in Syria, which marked the transition to agriculture. Tokens occur in the sixth millennium BC in China and in the fifth and fourth millennium in Europe and Africa, where they also coincided with the beginning of agriculture. The need for counting and record keeping

therefore may be attributed to farming, and in particular to the economy of redistribution typical of the early agricultural settlements.

It is important to understand that counting - the ability to determine the number of items in a collection – changed the economy. Counting and counters gave power to impose contributions and enforce their delivery. In other words, they gave control over the production and exchange of real goods. As I have discussed elsewhere, it is likely that the Near Eastern prehistoric tokens served for the administration of goods collected from communities on the occasion of seasonal festivals.[17] The created surplus of staple goods, such as grain and animal on the hoof, was the fulcrum of the redistribution economy and tokens played a key role in its administration.

It is remarkable that, within and without the Near East, the tokens were based on the same fundamental principle. Namely, the counters stood for real goods and by doing so abstracted commodities from reality. The heaviest and bulkiest loads of grain, or the most unruly animals, could easily be counted with the miniature counters. Moreover, the tokens made it possible to count goods, whether in the field or harvested, whether stored or promised. Collectively, the Neolithic communities of the core and periphery solved the same concern by the same symbolic means.

It is even more intriguing that geometric shapes were always selected to symbolize the merchandise. To each token shape was assigned a discreet meaning: in the Near East, a cone was a small measure of grain, and a sphere stood for a large measure of grain. In fact, it is likely that, when there was borrowing, the tokens were adopted wholesale, keeping the

[17] Denise Schmandt-Besserat, “Feasting in the Ancient Near East,” in Michael Dietler and Brian Hayden, *Feasts*, Smithsonian Institution Press, Washington, 2001, P. 391-403

same meaning attached to the same form. The similarity in the use of symbolism and the overall propensity for geometric shapes may point to a common human innate cognitive capacity.

It is also significant that, wherever they were adopted, the tokens were based on a same archaic technique of counting. As it is well documented by the numbers of counters held in the Near Eastern envelopes, the tokens were used in one-to-one correspondence. Six small measures of grain were shown with six cones, and ten small measures of grain with ten cones. Furthermore, tokens in multiple shapes are characteristic of a method of counting referred to as concrete counting. Concrete counting is characterized by using special number words, or special counters, to compute each particular category of items. For instance, small measures of grain could only be counted with cones and vice versa spheres could only be used to tally large measures of grain. Accordingly, the ubiquity of tokens demonstrates that concrete counting was the norm in many cultures. In other words, numeracy seems to have evolved similarly in many parts of the world, with a stage of archaic counting, such as concrete counting, preceding the acquisition of abstract numbers.

The tokens from the Near East and the peripheral regions have an unmistakable family resemblance in material, manufacture and size. They were used consistently for the same purpose, which is to be expected since they stem from the same origin. In comparison the distant group has a distinctly different feel. These tokens are larger, bulkier, and less carefully manufactured. The repertory of shapes is limited to cones, cylinders and mostly disks except at Jebel Moya where the collection has a rich number of original shapes never seen elsewhere. Both Vivara and Teotohuacan (of course independently) create disk subtypes by cutting them in halves,

thirds, quarters, which never happens in the core and periphery. Finally, some of the Teotihuacan tokens are cut in mica.

Surprisingly the tokens from the third, distant group, share some of the main features of the original Near Eastern tokens. All were mostly made of clay; they symbolized units of goods with geometric shape (Linda Manzanilla interprets the disks and their various subtypes as different quantities of tortillas); they served for concrete counting and accounting in one-to-one correspondence. The fact that many cultures, far or near, and in diverse time periods could readily borrow from one another, or invent independently, a similar system of counters of geometric shapes to refer to specific quantities of merchandise may point to a commonality of cognitive aptitudes in the preliterate world.

## Differences

The Near Eastern token assemblages distinguish themselves from all the others by being larger, and most importantly, by evolving from plain to complex forms. Where the excavators were vigilant, tokens were recovered by the hundreds in Mesopotamia and Persia. For instance, the site of Jarmo, Iraq, produced 1153 spheres, 206 disks and 106 cones. In comparison, the tokens in the distant group collections are also quite numerous, but those of the periphery are fewer.

The most significant difference between the tokens within and without the Near East is the fact that the former evolved, but the Chinese, European, African and Mesoamerican tokens did not. In Mesopotamia, after a period of about four thousand years when tokens showed little or no change, abruptly about 3300BC, the assemblage of the city of Uruk became very complex. It featured some seven additional types of tokens,

among which such new forms as triangles, quadrangles and parabollae, but also some more naturalistic shapes such as miniature heads of animals, vessels and tools. These main types were further differentiated into some 300 subtypes by the addition of lines or dots. The phenomenon of complex tokens extended beyond the metropolis to all the cities under Uruk's sway. Namely, the exact same types and subtypes of complex tokens have been recovered in Tello in Iraq; Susa and Chogha Mish in Iran and; Habuba Kabira and Jebel Aruda in Syria. No such complex assemblage is known outside the Uruk realm.

The multiplication of token shapes occurred simultaneously with the rise of the Uruk city state, which implies that the state administration was strengthened in order to control the production and movement of a larger variety of goods with greater specificity. This further underscores that, in the Near East, the need for counting and accounting went hand in hand with the growth of formal leadership.

## Tokens and Writing

The evolution from plain to complex tokens in Uruk is important because it started the series of events that led to the invention of writing. First, the Uruk accountants invented envelopes in the shape of hollow clay balls to hold together tokens representing debts. The second step was to mark, on the outside of the envelopes, the shape of the tokens held inside. This was done by impressing the tokens in the soft clay of the surface of the envelope. These impressions indicated the shapes or types of the tokens held inside the envelopes, while their number was shown in one-to-one correspondence. The hollow envelopes holding tokens did not have a long duration, but the markings made by impressing tokens on their face

survived. With a third invention, they were imprinted on solid lumps of clay – tablets. In time the impressed markings were complemented by more legible signs, still in the form of tokens, but traced with a stylus. Fourth, the final momentous invention was to confer a phonetic value to signs. The phonograms were no longer in the shape of tokens. The new signs were incised sketches of things easy to draw that stood for the sound of the word they evoked. The drawing of a man stood for the sound "lu" and that of the mouth for "*ka*," that were the sounds of the words for "man" and "mouth" in the Sumerian language. The phonograms were invented for composing individuals' names. These were written like a rebus, for example, the modern name "Lucas," could have been written with the two signs mentioned above "lu - ka." These archaic tablets are considered the origin of the cuneiform script.

Like the phenomenon of complex tokens, the metamorphosis of tokens into impressed markings, and the change from envelopes to tablets, occurred only in sites yielding a typical Uruk administrative assemblage that included special cylinder seals picturing the Uruk high priest, referred to as *En*, typical pottery bowls and jars, and a monumental architecture decorated with cone mosaics. The invention of writing can undoubtedly be attributed to the Uruk administration and probably to a single accountant, whose name has forever disappeared from human memory.

Things happened differently in China where forty five round clay envelopes, 4-6 cm in diameter, holding from one to seven small clay spheres inside were excavated at the site of Anhui Qianshan Xuejiagang and 60 more at Lintong Jiangzhai, in the Shanxi province.[18] The excavators report that many of the envelopes are marked with small holes and linear impressions but do not specify whether the number of markings

[18] Yan Zhi and Park Zaifu, "Ancient Counting tools from Excavations – From Shang Zhai and other Cultures" in Wang Yuxin, et allii, eds., *Beijing and Chinese Civilization*, Social Science Archive Publisher, Beijing 2006.

is related to the number of spheres inside. The resemblance of these artifacts with those from the Near East is remarkable, but the fact that the objects were found in graves, suggests a function different from those of Uruk. The Chinese envelopes with or without markings eventually disappeared without ever leading to any further development. In particular, the shapes and patterns of the large collection of pottery markings, for which the Neolithic Shuangdun culture is famous, show no likeness with the form of the contemporaneous tokens or their markings.[19] Furthermore, since the pottery markings of Shuangdun are not considered to be a script, the Chinese tokens can in no way be regarded as an antecedent of writing in China. The hollow clay envelopes used to hold tokens, that led to a dead end in China, but sparked off writing at Uruk, may be the best illustration of the serendipity that presided over the invention of writing.

Uruk invented writing and, by doing so, smote the first breach into the prehistoric Near Eastern token system. The new technology was in total contrast to the simple geometric counters. The token system had spread so swiftly and easily from culture to culture that in 7500 BC they were used in Tell Mureybet in Syria, as well as at Tepe Asiab in Persia, at the other end of the Fertile Crescent, The same clay counters remained for millennia a common feature in villages and towns of the entire Near East and beyond. In fact the 1500 BC Nuzi envelope filled with counters is prime evidence that, in the Near East, tokens served the illiterate long into historical times. Instead, writing proved hard to learn because it required a special eye-hand coordination to trace the signs; the two-dimensional script, which could not be manipulated by hand, was far more abstract and thus more difficult to use. Consequently, for several centuries writing remained

[19] Kan Xu Hang, Zhou Qun and Xu Da Li, eds., *Bengbu Shuangdun – A Report on the Neolithic Site*, Vols. 1-2, Anhui Provincial Institute of Antiquity and Archaeology, Bengbu Museum, Science Press, Beijing 2008.

confined to the administration of the Uruk city state, where only a score of Mesopotamian accountants could master it. Once writing became phonetic it became language and culture specific. Whereas tokens had united a vast number of cultures, writing became a great divide.

## Conclusion

The tokens were counters. They were tools of the mind, and as such, can shed light on the cognitive skills of preliterate humans.[20] The similarity of all the known assemblages of tokens of several continents and throughout time suggests that the simple system of counters matched some fundamental aptitudes of the human mind. Among the common diagnostic traits features the predilection for small three dimensional artifacts that can be manipulated with the fingers; the propensity to select geometric shapes to facilitate memory; the method of one-to-one correspondence to perform additions and subtractions; finally, concrete counting, which allows one to count one item at a time. Today's abacus, the ultimate heir of the tokens, which still helps children all over the world to master the art of counting, testifies that the tokens tapped into some fundamental universal cognitive predispositions of the human mind.

## Abstract:

Over the last decades, excavators in various parts of the world, in particular China, Europe and Africa, have reported finding clay tokens similar to the Near Eastern counters, forerunners of the cuneiform script. In this paper I

[20] Lambros Malafouris, "Grasping the concept of number: How did the sapient mind move beyond approximation," in Iain Morley and Colin Renfrew, *The Archaeology of Measurement*, Cambridge University Press, 2010, P. 35-42.

will argue that the ubiquity of geometric counters for counting in one-to-one correspondence highlights 1) a fundamental aptitude of the human mind; 2) the unique contribution of the Mesopotamian Uruk state administration that developed the system into writing.

**Key Words**: Tokens, Counters, Uruk, Shuangdun Culture,Teotihuacan.

# Discussion: Tokens in China, Europe and Africa

Seung Il KANG
Hannam University, KOREA

It was in 2002 that I first came across Professor Schmandt-Besserat's book *How Writing Came About*. At that time, I was a first year graduate student, who just came to the United States, and was struggling with the inundation of books and articles to read. Though not an Assyriologist myself, I was forced to take a mandatory course titled "Mesopotamian History Seminar" with Professor Jerry Cooper at the Johns Hopkins University. Some 100 page excerpt from *How Writing Came About* (its scholarly version is her two-volume work *Before Writing*) was included in the reading material for the course. I was not sure whether Professor Cooper had received a copyright agreement from Professor Schmandt-Besserat for distributing a part of her book to the students, but it was not my concern. My first impression was: "Denise Schmandt-Besserat? Gosh, it's a long name. I will be having trouble to remember her name when preparing for the final exam." But I never expected to meet her in person here in Korea, and to be a discussant of her paper at that.

It is a great honor and pleasure for me to be here today with such an eminent scholar as Professor Schmandt-Besserat, who is a world-renowned expert on ancient tokens and the origins of writing. Though I do not recall

much of her argument in *How Writing Came About*, I still remember that her theory that writing was developed from accounting system based on clay tokens dumbfounded me; with no previous knowledge of Sumerian writing system, I had this vague idea that writing might have been invented for religious purpose or for literature.

In this paper "Tokens in China, Europe and Africa – the Significance," Professor Schmandt-Besserat examines various types of tokens found throughout the world. She classifies them into three groups: (1) core – the ancient Near East, (2) peripheral – China and Europe (including Egypt and Central Africa) and (3) distant – Mesoamerica. She finds similarities among these tokens from various continents of the world in their geometric shapes and symbolism. She ascribes the similar features to "a commonality of cognitive aptitudes in the preliterate world." In the meantime, there are differences: the tokens in the ancient Near East evolved from plain to complex tokens to record a variety of manufactured goods as the city state of Uruk grew in size, whereas tokens in other places did not. She finally argues that the tokens in Mesopotamia eventually developed into Sumerian writing.

Now, Professor Schmandt-Besserat, could you please be kind enough to answer the following questions, so we can fully appreciate the significance of your research?

1. The archaeological context in which artifacts are found is crucial to understand how they were used and what they represented. Please explain the archaeological context of the clay tokens.
2. What are various types of counters or record keeping devices other than clay tokens that were used in preliterate cultures?
3. How many tokens have been discovered in Mesopotamia, and how do you identify the products or goods each token represents? Do you

find any rules governing which tokens label which goods (e.g., resemblance)? Or were most of the token shapes arbitrarily chosen or created?

4. Your theory that Mesopotamian tokens were the precursor to the Sumerian script is now widely accepted. Do you believe that the tokens account for the origins of all Sumerian signs? Some of the Sumerian signs were apparently direct heirs of tokens; but it appears that a lot of signs have no counterpart of tokens. And if this is the case, it may be seen as a weakness of your theory. It seems that you provide a partial answer to this question by stating, “the final momentous invention was to confer a phonetic value to signs. The phonograms were no longer in the shape of tokens. The new signs were incised sketches of things easy to draw that stood for the sound of the word they evoked.” Yet, one might ask whether this process of conferring a phonetic value to these newly created signs is to be seen as a continuation of the development from tokens, or an independent development of a writing principle accomplished without recourse to the token system.
5. In this paper, you show that the token system evolved into the invention of writing in Mesopotamia, but this was not the case in China. What factors do you think contributed to the difference?
6. Some people believe that Chinese writing may have been influenced or inspired by the Sumerian writing system through cultural diffusion. We would appreciate your comments on this.

# Khitan: Understanding the language behind the scripts

Juha A. JANHUNEN
University of Helsinki, FINLAND

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 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 the 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 (大文 *Dawen*) and the Khitan Small Script (小文 *Xiaowen*). 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]

[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

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

language are also detached from the graphic reality and cannot be considered as confirmed, as was already pointed out by Kwanten (1984, 1988).

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, 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 foreword starting with the early 1920s.[6] A

[3] The principal works on the Jurchen Script are Grube (1896), Kiyose (1978), Kane (1989), Jin Qicong (1984), and Pevnov (2004). Of these, only Kane comments also on the Khitan Large Script.

[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] Most research on the Khitan Small Script has mainly 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).

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 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.

---

[7] Qinggeertai & 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).

[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).

## 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 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

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

[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

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]

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

---

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 collection of Khitan Small Script "texts" in digital form was compiled by a Russian team in the 1960s and '70s, cf. Starikov & 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. 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).

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 Khitan. All other extant and historical Mongolic languages represent a single genetic node, corresponding to the language of the historical

[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.

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 which 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.

[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).

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 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]

[19] The Tungusic family also has two transitional (or mixed) branches. For a more detailed taxonomy of Tungusic, cf. Janhunen (2011).

[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 (2000).

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 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

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

compared with Ewenki-Ewen. 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 *).[22]

## 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

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

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.[23]

On the Mongolic side, the proximal pronoun has the oblique stem **e-xü/n-* > **üün-*, which has been compared with the possibly pronominal Khitan monosyllable 夭 *ún* †*un*. This comparison remains unverified, but it would seem to correlate with Khitan 夭爻 *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

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

surprising that the Khitan distal pronoun is 𘕕 *qi* 'that' : plural 𘕕𘐈 *qi.t*, suggesting an item totally different from Mongolic **te-*.[24]

*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 𘍞𘕿 *ci.ur-* †*jur-*, 3 𘐢𘕿 *hu.ur-* †*hur-*, 4 𘐈𘕿 *t.ur-* †*dur-*, 5 𘍿 *tau* †*tau*, 7 𘐪𘐴 $da_2$*.lo-* †*dal-*. The similarity to Mongolic 2 **jir-* (or **jïr-*),[25] 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 𘍠 *jau*, 1000 𘍢 *ming*, 10 000 𘐴 *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 𘍜 *is* †*is-*. This item is often compared with the Modern Mongolian 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

[24] In this case, the discrepancy has also been explained by drawing the Romanization of the Khitan character into question. Most recently, Wu Yingzhe (2011: 00-00) 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).

[25] 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).

Proto-Mongolic lineage, as is also confirmed by data from the "teens" in Jurchen-Manchu.[26] 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*, as 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.[27]

*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 (***): 𘠷𘖵 *heu.úr* †*haur* = **kabur* 'spring', 𘍿𘊱 $ju_3$*.un* †*jun* = **jun* 'summer', 𘈦𘍿𘖵 *n.am.úr* †*namur* = **namur* 'autumn',

[26] On the teens in Jurchen-Manchu, cf. Janhunen (1993). Athough 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.

[27] 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).

𐲖𐲢 *u.ul* †*uul* = **öbül* ‘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 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.[28]

*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 𐲑 *ai* ‘father’, 𐲒 *ia* ‘elder brother’, and 𐲓 *au* ‘elder sister’, have no obvious cognates on the Mongolic side. In other cases the correspondences are uncertain: 𐲔 *mo* ‘woman, mother, wife’ (plural 𐲔𐲕 *mo.t*) and 𐲖 *bu* ‘grandfather’ (in 𐲖 𐲑 *bu ai* id.) have been compared with Mongolic **eme* ‘woman’ resp. **ebü-gen*

[28] 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.

'old man', but the issue cannot be considered as fully concluded.[29] The comparison between 𘮩𘮆 *b.qo* 'child' (plural 𘮩𘮘𘮃 *b.hu.án*) and Mongolic **baga* 'small' > 'child' is also problematic.

There are, however, also good and non-trivial correspondences. These include 𘮃𘮃 *na.ha* †*naha* (plural 𘮃𘮃𘮏𘮍 *na.ha.án.er*) 'maternal uncle' = **naga-cu* id., 𘮓𘮁 *k.iú* ~ 𘮁𘮁 *x.iú* †*keu* 'younger sister' = **keii* 'child', and very probably 𘮆 *deu* †*deu* (genitive 𘮆𘮁 *deu.un*) 'younger brother' = **de*(*x*)*ii* id.[30] In two items, Khitan would seem to have a monosyllabic root (CV) where Mongolic has a bisyllable, apparently a derivative: 𘮀 *ku* †*ku* 'person' (genitive 𘮀𘮁𘮄 *ku.û.un*) = **kü-xün* id.,[31] 𘮏𘮘 *n.o* †*no* 'spouse' = **nö-kür* id., 'friend'. In the item 𘮏𘮊𘮋 *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

[29] 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.

[30] 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.

[31] Mongolic **kürün* 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).

𗼻𗴲 *go.er* ‘tent’ > ‘household’ = **ger*.[32] 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: 𘠋𘟂𘠎 *ci.i.is* †*cis* ‘blood’ > ‘blood relative’ = **ci-s-ü/n* ‘blood’, presupposing the monosyllabic root **ci*. As may be seen, this item 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.[33]

## 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

[32] 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.

[33] 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).

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 (1053-1175), 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.[34] 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 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.[35]

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

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

[35] For a general discussion of the areal and typological background of vowel rotation in the languages of Northeast Asia, cf. Janhunen (1981).

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', 孕 *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*).[36] 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 列化 *hu.ur-* †*hur-* = **gur-* vs, 4 令化 *t.ur-* †*dur-* = **dör-*.[37]

*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 the region, including Turkic, Mongolic, Tungusic, and even Japonic.[38] 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.[39]

---

[36] 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.

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

[38] 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.

[39] 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.

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* = 丵 *po* 'time' = Pre-Proto-Mongolic **pon* (= Jurchen †*po*) > Proto-Mongolic **xon* > Modern Mongolian *on* 'year'.[40] Today, we have a few more examples, including 丵 木 *p.ar* †*par*(*a*) 'people' = **para* > **xara/n* : (plural) **xara-d* > *ar*(*a*)*d* id., 丵 平 夂 *p.ul.uh* †*pulu*(*-h*) 'intercalary month' < 'more' = **püle-xü* 'more, extra', as well as 丵 夾 令 *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,[41] while in the 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 几 *ku* †*ku* 'person' (genitive 几 爪 厶 *ku.û.un*) = **kü-xün*

[40] 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.

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

id. The variation in 𐲓𐲚 *k.iú* ~ 𐲍𐲚 *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 𐲍𐲚𐲉 *x.ui.ri* 'to arrive', which may or may not correspond to Proto-Mongolic **kür-* id.[42] Before original back vowels, spirantization is confirmed by 𐲁𐲗 *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 𐲖𐲉 *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".[43] 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 𐲘 *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 𐲛 *tau* †*tau* = **tabu/n*, 𐲁𐲗 *heu.úr* †*haur* = **kabur* 'spring'. The fact that there is no medial consonant in these cases is confirmed by the use of the characters 𐲘 *jau* and

[42] 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-*.

[43] 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.

𘰾 *tau* for the Chinese term 𘰽𘰾 *jau.tau* †*jautau* 'punitive commissioner' = 招討 **jawtaw* (*zhaotao*). The use of the character 𘰾 *tau* in the word 𘰾𘱁𘰭 *tau.li.a* 'hare' = **taula-i* id. is also significant, though the phonological background of this "rebus" remains unclear.[44]

The fate of **x* in monophthongoid contexts in Khitan is less clear, which is why comparisons like 𘰿 *ún* †*un* = ?**e-xü/n-* are uncertain. A possible example would be offered by 𘰻 *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 𘰻𘰼𘱀 *qa.ha.án* 'of the emperor' is also enigmatic since it seems to contain a medial **g*, as in 𘰼𘰺 *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 Koreanic, while other families, including Mongolic and Tungusic, show a less systematic, or at least chronologically later, tendency to lose vowels.[45] A special feature of

[44] 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).

[45] 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

Jurchen-Manchu (and certain forms of Mongolic) is the loss of vowels in open second syllables (the so-called "Mittelsilbenschwund").

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 𘁆𘁏 *b.as* 'again, also' = **basa* id. and 𘁓𘁔 *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, 𘁅𘁇 *i.ri* 'name' = **nere* id. (with a secondary initial nasal) and 𘁌𘁇 *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.[46] 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 𘁐𘁃𘁍 *t.qo.a* 'chicken' = **takï-xa* id., 𘁑𘁄𘁋 *mu.ho.o* 'snake' = **moga-i* id., and also 𘁒𘁂𘁍 *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.

---

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.

[46] 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.

*

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 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 similarities are conspicuous enough to have been evident even for the naïve 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

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

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

——(1992), Mongolica im Alttürkischen, pp. 30-56 in: *Bruno Lewin zu Ehren* 3. Bochum.

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

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.

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

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

——(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.

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

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

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

——(2011), The Expansion of Tungusic as an Ethnic and Linguistic Process, pp. 000-000 in: *Advances in Tungusic Studies*, Lindsay Whaley & Andrej Malchukov (eds.), Tunguso-Sibirica 00. Wiesbaden: Otto Harrassowitz (forthcoming).

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.

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

——(2006), Khitan and Jurchen, pp. 121-132 in: Alessandra Pozzi & Juha Janhunen & Michael Weiers (eds.), *Tumen Jalafun Jecen Akū: Manchu Studies in Honour of Giovanni Stary*, Tunguso-Sibirica 20. Wiesbaden: Harrassowitz Verlag.

——(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).

——(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.

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

——(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, pp. 265-308 in: *Mongolian Studies*, Louis Ligeti (ed.), Bibliotheca Orientalis Hungarica 14. Budapest.

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".

Qinggeertai 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.

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

Rybatzki, Volker (2003), Middle Mongol, pp. 57-82 in: *The Mongolic Languages*, Juha Janhunen (ed.), Routledge Language Family Series 5. Routledge: London & New York.

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 & M. Probst (1970), *Materialy po deshifrovke kidan'skogo pis'ma* 1-2. Moskva: Akademiia Nauk SSSR.

Sun Bojun 孙伯君 & 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 武内康則 (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: Kyouto 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 Gorō (1964), An Analysis of the Major Ch'i-tan Characters, *Memoirs of the Research Department of the Toyo Bunko* 23: 119-135.

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

——(2000), A Modest Proposal on the Decipherment of the Khitan-Chinese Bilingual Text of the 1134 (the Langjun Inscription), Paper presented at the 43rd Annual Meeting of the Permanent International Altaistic Conference (Chateau Pietersheim, Belgium, September 3-8, 2000).

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

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.

——(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.

——(2009), A Brief Discussion on the Vowel Attachment in the Khitan Small Script, *Voprosy filologii, Seriia “Uralo-altaiskie issledovaniia”* 1: 26-30.

——(2011), Deciphering Some Demonstrative Pronouns in Khitan Small Script, *Altai Hakpo* 00: 000-000 (forthcoming).

—— & 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.

## Abstract:

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

# Discussion: Khitan: Understanding the language behind the scripts

LEE Seong Gyu
Dankook University, KOREA

## 1. General Review

Professor Janhunen of the University of Helsinki is both a well-known Altaicist and also a leading scholar in the research of Manchu and Mongolian. The paper presented at this conference relates to his previous research on the tribes and peoples of the Manchu region. In particular Janhunen recently attempts extensive analysis of not only the languages themselves but also their relationship with history, culture and folk customs.

By attempting the phonological analysis of Khitan, his work can be seen as very original. One reason for this being that the number of scholars referencing the phonology of Khitan in detail is very small. Also, another characteristic of the work is that Prof. Janhunen has described Khitan as being located beside Mongolian and not as its direct ancestor as many scholars have before. More specifically he would like to place Khitan on a separate branch of Altaic having equal standing with Turkic and Mongolic.

On the other hand, the Khitan small script has not yet been completely deciphered and many difficulties persist in the research. The need to

supplement the research to overcome these difficulties with pieces of literature and annotations written in Chinese is a splendid opinion.

## 2. Discussion

The presenter would like to locate Khitan not as a direct ancestor of Mongolian but rather as a distinct language. However many of the vocabulary items cannot be freed from the Mongolic category. Is this an inability to establish certainly the characteristics differentiating Khitan from Mongolic?

In the research of Sun Bojun 孫伯君 and NIE Hong-yin 聶鴻音 (2008) the authors restored some 300 Khitan words written with Chinese characters through the phonology of Middle Old Chinese and compared them with Altaic languages. In this research considerable reconstruction of Khitan's vowels and consonants has been made. Do you have any intention to use this research in the future?

If one compares Khitan written in Chinese with Korean, the majority of vocabulary as seen below can be linked to Korean (Yi Seongkyu 2010, 2011). Will you do research comparing Korean to Khitan in the future?

> (阿斯 / *as / 아사(달), 孤穩 / *güγön / 곱, 捏咿呪 / *neri / 날, 女古 / *nürgü/ 누른, 于越 / *üγö/ 우거, 樺 / *qa / 활, 討 / *taw / 다섯, 陶里 / *toil / 토끼, 爪 / *jo / 온, 阿主 / *ajü / 아주(머니), 暴里 / *böri / 부리, 墮瑰 / *dögü / 문(가야어), 剋 / *ga / 벼슬 이름(帥). 彌里 / *miri / 모라(牟羅).

If we look at the recent trend in Khitan research, very detailed analyses of the pronunciation of the Chinese characters used to write Khitan are taking place (Wu Yingzhe 吳英喆, Jiruhe 吉如何), but judging from this why is no detailed phonological research about Khitan vocabulary being done?

It is known that Prof. Janhunen has a wide understanding of Manchu-Tungus. To what degree are Khitan and the Manchu-tungusic languages related?