

RECONSTRUCTING THE LANGUAGE MAP OF PREHISTORICAL NORTHEAST ASIA

Juha Janbunen

Focusing on the northern part of the East Asian region, the paper discusses the principles of positioning languages and ethnic groups in time and place.¹ The main argument is that ethnic groups can be followed back in time only on the basis of the genetic lineages of their languages. In Northeast Asia, eight well-documented lineages of languages can be established, including Sinitic, Turkic, Mongolic, Tungusic, Amuric, Koreanic, Japonic, and Ainuic. For each of these lineages, it is possible to postulate actual proto- and prehistorical political, cultural, and territorial connections. The modern linguistic map of the region is a result of relatively recent expansions. These expansions have not necessarily involved large-scale human migrations. Rather, they are results of linguistic diffusion, the basic mechanism of which has been language shift. The impact of language shift is also visible in the synchronic areal and typological patterns exhibited by the languages of the region.

THEORETICAL PREREQUISITES

Every natural language is supported by a speech community, that is, by a group of people speaking the language concerned. A basic property of any speech community is that it has a geographical location which corresponds to the territory occupied by its speakers. The territories covered by speech communities are typically dependent on geographical factors, such as orographic and hydrographic features, as well as climate and vegetation. The population size of a speech community and the size of the territory it occupies are also governed by cultural and political factors. Ultimately, it is historical chance that determines which speech community occupies any given physical region, that is, what language is spoken in that region.

1 This is a slightly revised and updated version of a paper first presented at Asia House, London, on 30 October 2006, as part of the Asian History Lecture Series supported by Sungjoo Group. The author thanks Professor Gina Barnes, the official discussant, for her valuable comments on the paper.

Speech communities evolve over time. This evolution concerns both the demographic profiles and the geographical locations of speech communities. The three principal processes that can affect the link between a speech community and its geographical location are expansion, contraction, and translocation. In this context, expansion means that the territory of the speech community becomes larger, often due to cultural innovation, demographic growth, and increasing political power. Correspondingly, contraction means that the territory of the speech community becomes smaller, often in connection with the expansion of neighbouring speech communities. Translocation, finally, means that the speech community as a whole shifts its location from one territory to another, a development that can take place due to environmental, cultural, or political reasons.

Speech communities are best defined by their boundaries. A straightforward criterion for determining whether people speak the same language as their neighbours is to see whether they can communicate with each other linguistically. Whenever there is an abrupt discontinuity of intelligibility there is also a linguistic boundary that marks the limits of separate speech communities. The only significant modification that this definition needs is connected with the phenomenon of bilingualism. Bilingual individuals are members of two speech communities at the same time, and they can therefore communicate with members of both communities. This means simply that the territories of speech communities can overlap on their margins. This overlapping is never complete, however, and each speech community typically has a core area dominated by monolingual individuals.

Linguistic boundaries are formed diachronically by the process of divergence, that is, by the gradual diversification of an originally uniform (for purposes of communication) language into two or more distinct idioms. Linguistic diversification is often the result of expansion, for increasing geographical distance tends to favour increasing linguistic distance and diversity as well. It goes without saying that mutual intelligibility, when viewed in the context of linguistic divergence, is a transitional phenomenon, for related languages retain some degree of intelligibility even after their separation. There are also examples of dialect chains in which every single local form of speech is intelligible with its immediate neighbours, while the extremities of the chain are mutually unintelligible. This situation does not alter the fact that even dialect chains are ultimately delimited by sharp linguistic boundaries against their external (unrelated) neighbours.

It is not accidental that the phenomenon of linguistic divergence has become the most important tool of ethnic history. Typically, a speech community corresponds to an ethnic group, that is, a population which in addition to its language shares a distinct complex of cultural patterns and biological heritage. Although there are many exceptions, involving cases of multicultural and multiracial societies

speaking a single language, the basic pattern all over the world is nevertheless that linguistic unity tends to correlate with cultural and biological coherence. Consequently, linguistic boundaries also involve cultural and genetic boundaries, and ethnic identities on both sides of a linguistic boundary are conceived of as being based on a combination of linguistic, cultural, and biological criteria.

The correlation between language, culture, and genes is, however, a synchronic phenomenon. The further back in history we go, the more diversified the sources of any given modern ethnic group are. In particular, it is virtually impossible to trace back an ethnic group on the basis of its cultural or biological heritage alone. The only ethnic marker that can unambiguously be followed backwards in time is the linguistic lineage, that is, the genetic (in the linguistic sense) identity of the language. This is the basic principle of comparative linguistics, and thanks to comparative work we have today a relatively comprehensive understanding of the language families of the world. Diachronically, each language family corresponds to one or more stages of past linguistic expansion, accompanied by diversification.

It is important to realize that linguistic continuity in time does not imply continuity in place. Although we can follow any given linguistic lineage backwards in time to the corresponding protolanguage that can be reconstructed on the basis of the relevant comparative evidence, we have in general no direct information on the original location of the protolanguage-level speech community. Recent progress in the dating methods of both archaeology and, in particular, human genetics, suggests that cultural features and genetic markers often exhibit a considerable local continuity. This continuity does not, however, extend to languages. Evidence from historically documented parts of the world, such as Europe, suggests that languages are often changed where populations and cultures remain stable. Of all factors relevant to ethnic identity, language is the least stable locally.

Divergence in linguistic evolution is to some extent balanced by the opposite phenomenon of convergence, by which is meant the tendency of languages to influence each other. When two or more mutually unrelated languages interact in the context of a geographically or culturally coherent region, they transmit linguistic features, such as structural properties and lexical items, across linguistic boundaries. As a result, the interacting languages become more similar to each other. The similarities originating from language contact are, however, secondary as compared with the primary similarities due to genetic relationship, and the comparative method allows a distinction to be made between the two types of similarities.

Although convergence inherently increases the mutual intelligibility of the interacting languages, it can never make two idioms so similar that they merge into a single language. Cases in which two speech communities actually merge do not involve linguistic convergence but language shift, that is, the loss of one

of the languages in favour of the other. Language shift is probably the most important mechanism of ethnic history, for a change in the linguistic identity of a population normally has profound long-term implications for the cultural and political orientation of the population concerned. It may be noted that language shift can also take place between related languages, including closely related languages that still retain some degree of mutual intelligibility. In such cases, language shift results in decreasing diversity within the expansive language family.

Contact features in languages are conventionally divided into several categories depending on how the underlying linguistic interaction takes place and how it is manifested in the linguistic substance. The most fundamental distinction should probably be drawn between processes that merely modify the linguistic substance of one or more interacting languages, and processes that involve language shift and language loss. Examples of the former type of process are *adstrata* (exchange of occasional features between adjacent languages) and language unions (systematic levelling of structural features in the context of intensive interaction). Examples of the latter type of process are *substrata* (absorption of non-dominant local languages), *superstrata* (absorption of dominant non-local languages), and *creolization* (abrupt language shift in favour of a new dominant language).

THE LINGUISTIC TAXONOMY

In the widest sense, Northeast Asia as a geographical and ethnohistorical region can be defined as the entire northeastern part of the Eurasian continent, delimited by the Yenisei in the west and the Yellow River in the south. In the northeast, the region extends, in principle, to the Bering Strait. In a somewhat narrower framework, Northeast Asia may be defined as comprising the territory between the Amur and Yellow River basins, including the Korean Peninsula and the Japanese Islands in the Pacific coastal zone, but excluding the northeasternmost limits of what is today the Russian Far East. Focussing on this narrow definition, and not counting recently introduced colonial languages, notably Russian, the languages today spoken in the region can be divided into eight lineages, each of which represents, in light of current knowledge, a separate language family:

(1) Sinitic, comprising the various forms of Chinese spoken all over Northern China and Manchuria. All of these forms of Chinese belong to the context of Mandarin (or Northern Chinese), as opposed to the other Sinitic languages spoken mainly south of the Yellow River basin. In spite of considerable regional variation, which often exceeds the limit of mutual intelligibility, Mandarin is a relatively uniform complex of Chinese-based idioms. Due to the presence of transitional dialects, it may even be viewed as a single dialect chain, although there are also several highly aberrant local forms that are probably best classified

as separate languages. The general coherence of Mandarin suggests that its current territorial extension is due to a relatively recent expansion from a limited source region.

(2) Turkic, spoken across the entire Central Eurasian belt from Anatolia and Eastern Europe to Siberia and Eastern Turkestan. The territorial weight of Turkic speakers today is concentrated in the west, where the language family is divided into two primary branches known as Bulghar Turkic (represented by the Chuvash language on the Volga) and Common Turkic (represented by all other extant Turkic languages). Common Turkic, in particular, is a conspicuously uniform group of languages, suggesting a relatively recent breakup of the protolanguage. All of the Turkic languages spoken in Northeast Asia today belong to the Common Turkic branch, but within this branch, the internal diversity grows towards the east and north, where three relatively clear-cut sub-branches are present in the Lena basin (the Yakut group or Lena Turkic), the Sayan region (the Tuva group or Sayan Turkic), and the Upper Yenisei basin (the Khakas group or Yenisei Turkic). Two other aberrant Common Turkic languages (Salar and Yellow Uighur) are spoken in the Upper Yellow River region (the Gansu-Qinghai provinces of China).

(3) Mongolic, distributed mainly in Mongolia (including Inner Mongolia), but extending also to the adjacent parts of Manchuria (the Amur, Nonni, and Liao basins), Siberia (the Baikal region), Sinkiang (Jungaria, the Tianshan region), and Northern Tibet (the Kuku Nor region). Most of the Mongolic languages retain some degree of mutual intelligibility and form the so-called Common Mongolic group, which also comprises a diaspora entity in the Caspian region (Volga Kalmuck). A distinct sub-branch with several mutually unintelligible languages, known as Shirongolic, is located in the Upper Yellow River region. As a relict feature, a Mongolic language (Moghol) has also been spoken until recently in Afghanistan.

(4) Tungusic, forming the territorially largest language family of Northeast Asia and covering most of Siberia between the Baikal region in the south, the Arctic coast in the north, the Yenisei basin in the west, and the Pacific coast (including parts of Kamchatka and Sakhalin) in the east. Most of the Tungusic territory is occupied by the internally fairly uniform Northern Tungusic (Ewenki-Ewen) branch, but more linguistic diversity is found in Manchuria (the Amur, Sungari, and Ussuri basins), where conventionally two other branches, Amur Tungusic and Southern Tungusic (Jurchen-Manchu) are distinguished. There are also some secondary diaspora entities in Sinkiang (Jungaria), dating back to the colonial period.

(5) Amuric, synchronically represented by a single language, Ghilyak (also known as Nivkh), spoken in the Amur Delta region and on Sakhalin. Since Ghilyak has no known linguistic relatives, it may be viewed as a genetic isolate. Within

the language there are, however, relatively large internal dialectal differences, allowing three distinct main dialects to be distinguished (Amur Ghilyak, North Sakhalin Ghilyak, and South Sakhalin Ghilyak).

(6) Koreanic, spoken very consistently within the physical borders of the Korean peninsula, a geographical entity where in recent historical times no other indigenous languages are known to have been spoken. Like Amuric, Koreanic may be viewed as a genetic isolate, represented only by a single language, Korean. There are, however, dialectal differences which occasionally, especially in marginal areas (such as Jeju Island), exceed the limit of mutual intelligibility.

(7) Japonic, comprising two distinct languages: Japanese (proper) on the Japanese Islands, and Ryukyuan on the Ryukyu Islands. Both Japanese and Ryukyuan involve dialect chains, with faraway members of the chain (such as the Japanese dialects of Southern Kyushu and Northern Honshu) being mutually unintelligible. The internal diversity is probably greater within Ryukyuan, a circumstance that could possibly allow this entity to be viewed also as a group of several closely-related but separate languages. As a whole, Japonic may be viewed as a small language family, since it has no known external genetic relatives. Unlike Koreanic and Amuric, Japonic is not a genetic isolate in the strict sense.

(8) Ainuic, comprising the Ainu language spoken on the Japanese island of Hokkaido (Ezo) as well as, historically, on Sakhalin and the Kuril Islands. Very much like Amuric, Ainuic is a group of three relatively distinct main dialects, but is nevertheless probably best classified as a single language. Like Amuric and Koreanic, Ainuic is a genetic isolate with no known genetic connections.

Looking at these eight language families of Northeast Asia, it is easy to see that some of them continue to be expansive up to the present day, while others have been rapidly regressing in recent history. The most expansive family has been Sinitic, which today has well over one billion speakers, most of whom speak various forms of Mandarin. Most importantly, in pre-modern times (starting about 1870), Mandarin Chinese spread with an increasing intensity to Manchuria (the so-called 'Three Eastern Provinces') and Inner Mongolia. This development, which belongs to the context of colonialism, has caused the regression of the local languages of Manchuria and Inner Mongolia. Before the colonial period, Chinese was only marginally present in the southernmost parts of Manchuria (Liaodong).

A second category among the language families of Northeast Asia is formed by Turkic, Koreanic, and Japonic, all of which have today between 70 and 120 million speakers. All three families have been expansive until recent times, though the most recent expansion of Koreanic has been territorially very limited (covering only the Yalu and Tumen river basins). Japonic, however, has made an important territorial conquest by the colonisation of Hokkaido (also starting about 1870), which had previously formed the principal territory of Ainuic (Hokkaido Ainu).

On Sakhalin and the Kuril Islands, the spread of Japanese as a colonial language was counterbalanced by the parallel expansion of Russian from Siberia.

A third category is formed by Mongolic and Tungusic, two families whose numbers of speakers were counted in millions until recent times. The combined effect of the Chinese colonisation of Manchuria and the Russian colonisation of Siberia has, however, led to the decline of all Tungusic languages, including, in particular, Manchu, the official language of the Manchu Empire of the Qing (1644–1911). As a result, Tungusic languages (mainly Ewenki and Manchu) are today spoken by less than 50,000 people. Mongolic, by contrast, survives in Mongolia, as well as among some diaspora populations (Shirongolic), so that there are still today well over five million people speaking various forms of Mongolic. Even so, Mongolic has not been expansive recently, and any growth in the number of speakers has been due to natural population growth, rather than territorial expansion.

Finally, Amuric and Ainuic have, in historical times, been systematically losing ground both territorially and in terms of numbers of speakers under the impact of not only the recent colonial languages (Chinese, Japanese, Russian), but also due to the expansion of languages representing other local families (Mongolic and Tungusic). The number of Ainuic speakers may have reached a historical height of some 100,000 people immediately before the colonial period, while the number of Amuric speakers seems to have been more or less stable at less than 5,000 people as far backwards as the situation can be followed. Today, Ainuic is virtually extinct, while Amuric may still have some hundreds of speakers.

It has to be noted that the modern numbers of speakers and the current demographic trends are of relatively little informative value when we try to approach the question concerning the earlier history and prehistory of the speech communities and language families concerned. The fact that some languages have had, or still have, an official status in political states is also of not much consequence for ethnohistorical conclusions. Of course, the fates of languages are intimately connected with their political status. It is not surprising to see that the most expansive language families in Northeast Asia (Sinitic, Turkic, Koreanic, Japonic) are represented by one or more modern state languages. Language families whose former political significance has diminished (Mongolic, Tungusic) have ceased to be expansive, while language families that have never had a political role in historical times survive today only in a rudimentary form. In fact, the current wave of global linguistic extinction is affecting not only a large number of individual languages but also three entire language families in Northeast Asia: Ainuic, Amuric, and Tungusic.

THE CHRONOLOGICAL DEPTH

Compared with the elaborate scientific methods of dating used in modern archaeology and human genetics, the dating of linguistic substance and the measuring of linguistic distances is still an undeveloped field. The methods of diachronic and comparative linguistics, including the method of internal reconstruction, allow remarkably reliable conclusions to be made about the relative dating of the chronological layers present in any actual language or language family. They also allow the primary (inherited) and secondary (borrowed) elements of languages to be separated with no ambiguity. In principle, these methods resemble the stratigraphic dating of archaeological material, which basically yields only relative datings. When it comes to absolute dating, however, linguistics faces many unsolved problems.

Ideally, absolute dating should be context-free. The only context-free method of linguistic dating available today is that of lexicostatistical glottochronology, which is based on the tenet that lexical items in any language, especially in the core part of the lexicon (basic vocabulary), are replaced at a constant and universally fixed rate. In spite of criticism presented against the method, its basic assumption is sound and corresponds to the empirical knowledge that languages change, but not so rapidly as to sever their communicative functions. Unfortunately, however, the speed of lexical replacement (and of linguistic evolution, in general) is not completely fixed, but varies within rather broad limits. For each particular case, the actual speed should be calculated by calibrating the basic formula, a procedure that so far appears impossible to accomplish. Therefore, in the absence of calculations, intuition based on experience still remains the best tool of the diachronic linguist.

There is also chronological information that is not context-free. Linguistic material, especially the lexicon, is inherently linked to a variety of contextual factors, including ethnographic details and archaeological stages, which can be dated. Since information on relative dating is readily available from any language, even a single point of reference with an absolute dating can allow an entire network of languages to be placed in a fairly accurate framework of absolute chronology. This is, essentially, the method of linguistic palaeontology, as understood in the chronological sense. The method can also yield information on the geographical location of the ancestral forms of modern languages, but, unfortunately, the geographical information thus obtained is often either too trivial or confusingly diffuse.

In Northeast Asia, with relatively early traditions of writing, we also have actual historical and textual information on the linguistic evolution of the region. In fact, there are two written languages more or less directly corresponding to the protolanguage level of linguistic reconstruction: Middle Mongol and Middle Korean (both around 12th to 15th cc.), which for all practical purposes are identical

with Proto-Mongolic and Proto-Koreanic, respectively. Two other written languages, Old Turkic and Old Japanese (both around 6th to 10th cc.), correspond chronologically to important secondary protolanguages: Proto-Common Turkic (without Bulghar Turkic) and Proto-Japanese (without Ryukyuan), respectively. The Tungusic written language of Jurchen (12th to 16th cc.) is less informative for purposes of absolute dating, though it does show that Jurchen (with Manchu) was already a distinct branch of Tungusic at the time when writing was introduced.

For some language families in the region, the available written information even predates the beginning of the modern linguistic diversity. All of the language families concerned must have already had some internal diversity at the time of the protolanguages, though in most cases this diversity has been lost without a trace. In the case of Mongolic, however, we have actual information on a group of languages that represent a branch collaterally related to Proto-Mongolic. This branch, termed Para-Mongolic, is represented by the well-documented but imperfectly understood Khitan written language (10th to 13th cc.). A similar case is probably involved in Old Korean (7th to 10th cc.), which may represent a lineage not fully identical with the one leading to Middle Korean and Modern Korean.

The most ancient written language in Northeast Asia is, of course, Chinese, but, unfortunately, the nature of the Chinese script, when used logographically, does not allow a detailed phonetic reconstruction of the earlier forms of the language. Even so, the picture based on a comparative analysis of the modern Sinitic languages ('Chinese dialects') is corroborated by conventional philological work on textual sources (the so-called rhyme tables) of the Middle Chinese period (7th to 10th cc.), suggesting that this is the approximate level of Proto-Sinitic. As a result, the common protoform of at least most of the extant forms of Chinese may be dated roughly to the historical period of the Tang dynasty (618–907 *az*). The preceding stages, known as Old Chinese, probably represent a variety of separate lineages more or less closely related to, but not identical with, the lineage of Middle Chinese.

It may be concluded that the sources of the extant diversity within most of the language families in Northeast Asia can actually be dated relatively accurately even in absolute terms. The only two language families for which no early written material is available are Amuric and Ainuic, but since both of them are genetic isolates (or almost isolates), the corresponding protolanguages must in any case be dated to relatively recent times. On the other hand, the pre-modern historical and ethnographic data on the Ghilyak and Ainu populations (18th to 19th cc.), including native sources of folkloric material and local history, suggest that the distribution of the extant dialectal groups dates back some centuries.

An important general circumstance evident from the historical information on the languages of Northeast Asia is that all language families in the region are diachronically relatively shallow. This circumstance is confirmed by the lexical

data, which for every language family concerned contain not only a coherent set of genetically diagnostic cognates, but also an extensive corpus of cultural vocabulary covering a variety of thematic fields pertaining to highly developed technologies and social structures. There is nevertheless some chronological layering, which allows the language families to be divided into four age groups:

(1) At the shallow end, there are the genetic isolates Amuric and Ainuic, which have a dialectal depth of perhaps no more than 500 years. In both cases, the origination of the extant dialectal diversity is connected with the geographical division of the speech communities into clearly separate territorial groups (Sakhalin and the Amur Delta region in the case of Amuric, and Hokkaido, Sakhalin, and the Kuril Islands in the case of Ainuic). We have no direct historical information as to the actual reasons for the expansion and territorial dispersal of the Amuric and Ainuic families, but whatever the underlying developments may have been, they are likely to have taken place between 500 and 1,500 years ago.

(2) Only slightly more distant in the past lie the origins of the Mongolic and Koreanic families, with the protolanguages datable to between 500 and 1,000 years ago. It is, however, interesting to note that linguistic differentiation has proceeded much further in Mongolic than in Koreanic. Mongolic today is a medium-size family with around 10 distinct languages, while Koreanic is represented by a single language. This is, without a doubt, connected with the difference in the historical and geographical contexts of the two language families. The history of Mongolic can with certainty be connected with the rise of the historical Mongols (11th to 12th cc.), with most of the geographically more distant Mongolic languages being relicts of the Mongolian Empire and its territorial conquests. The Korean Peninsula, by contrast, is a compact region where political unity has only increased with time. The expansion of Korean can be connected with the impact of the unified Silla state (668–936) and the consistent northward push of the northern border of Korea during the subsequent Goryeo (936–1392) and Joseon (1392–1910) dynasties.

(3) The historically documented correlation (rhyme tables) with the Tang dynasty places Proto-Sinitic some (but not many) centuries earlier than Proto-Mongolic and Proto-Koreanic, that is, to somewhere between 1,200 and 1,400 years before the present. There are indications that some of the extant Sinitic languages (especially the Min or Hokkien complex) may have separated from the context of the rest of Sinitic even slightly before the Tang period.

(4) A contemporary of Proto-Sinitic was Proto-Common Turkic, but Proto-Turkic, the ancestral form of Common Turkic and Bulghar Turkic, must lie deeper and may be dated to between 2,000 and 2,500 years ago. This is clearly indicated by the fact that the most important diagnostic feature of Bulghar Turkic, the phenomenon of the so-called rhotacism-lambdacism (involving the development of sibilants into liquids under certain conditions), is attested in

a large corpus of loanwords obviously borrowed before the Proto-Common Turkic period into an early form of Mongolic. A similar age can be postulated for Proto-Japonic and Proto-Tungusic: in these cases, a relatively early dating of the protolanguages is required by the historical levels established for Old Japanese and Jurchen, respectively. Another entity of a similar time level seems to have been the common ancestor of Mongolic and Para-Mongolic, though the dating in this case remains difficult to verify because of our insufficient understanding of the Para-Mongolic (Khitan) sources.

Summarizing the situation, the language families currently present in Northeast Asia have a depth ranging from 500 to 2,500 years. Many attempts have been made to project the established protolanguages further back in prehistory by looking for external relations that would allow the postulation of larger and deeper-level language families. In most cases, these attempts have been unsuccessful. A case in point is the so-called Altaic Hypothesis, which involves (in different versions) the assumption of a primary (genetic) relationship between up to five language families of the region: Turkic, Mongolic, Tungusic, Koreanic, and Japonic. There is a growing consensus today that the similarities on which the Altaic Hypothesis was based are due to secondary interaction (contact phenomena). Even proposals made about binary links between some of the Altaic entities, such as Koreanic and Japonic, or Mongolic and Tungusic, remain at the level of unverified hypotheses.

It may be asked why there are so many such shallow language families in Northeast Asia. When we look at other regions, especially in Eurasia, we see highly expansive language families with a considerable diachronic depth and geographical extension. Such language families are found both in Europe (Indo-European), the Near East (Semitic), and Southeast Asia (Austronesian). The probable explanation is that the expansion of languages is so closely connected with contextual factors, such as culture and geography. Northeast Asia has simply not been reached (before colonial times) by the large and deep language families of the continent. In this respect, Northeast Asia is more similar to North America.

There is, however, one large and deep language family that is present in Northeast Asia: Sino-Tibetan. Although still doubted by some diachronic linguists, the relationship between Sinitic and the rest of the so-called Sino-Tibetan languages seems impossible to deny. The internal structure of Sino-Tibetan, with many still unclassified branches, is still under discussion, but a recent suggestion that Sinitic is, indeed, relatively closely related with the Tibetan (Bodic) branch of the family appears well based. Even so, Sinitic is the only branch of Sino-Tibetan that is predominantly located in Northeast Asia, while the other branches are biased towards Southeast Asia. This circumstance has obvious implications for the original location and dating of the Sino-Tibetan protolanguage.

THE NETWORK OF HOMELANDS

Like all expansions, linguistic expansions always have an expansion centre. It follows from the principles of comparative linguistics that any protolanguage underlying the synchronic diversity of a language family must have been spoken in a compact region, occupied by an actual speech community. The location of the protolanguage may be identified as the linguistic homeland of the corresponding language family. The idea of a direct correlation between linguistic, cultural, and racial continuities has often led to the false assumption that a linguistic homeland is also a cultural and a racial homeland. This is not so, however, and the more we have learnt of archaeological dating and human genetics, the more obvious it has become that the movements of languages on the map are essentially independent of cultures and genes.

Since all the known protolanguages of Northeast Asia are relatively shallow and more or less contemporaneous with each other, it is possible to reconstruct parts of the earlier language map of the region by placing the protolanguages in their presumable homeland positions. The map we can thus create may be dated to about $1,500 \pm 1,000$ years before the present. It is important to stress that this time level is not prehistorical, but, rather, protohistorical. The protohistorical period in Northeast Asia is characterized by the presence of many kinds of written sources (both occasional inscriptions and actual histories), in which ethnic groups, cultures, and languages, are mentioned. These sources are, however, generally very diffuse, and reconstructing a language map on their basis alone would be impossible. A much more reliable source of information is provided by the available linguistic data.

There are many ways in which linguistic data can be used for establishing the locations of linguistic homelands. One celebrated principle is to look for the point of greatest genetic diversity within a language family. Since diagnostic differences (isoglosses) with a high relative age represent the chronologically earliest divisions of a language family, their location on the map is often indicative of the geographical origins of the family. Many language families have one relatively limited core region with a considerable genetic diversity, and one or several larger marginal regions with much less internal variation. Quite often there is independent information confirming that the more diversified core region also corresponds to the linguistic homeland of the family. In Northeast Asia, this is the case with Tungusic, which has a large and linguistically uniform northern extension in Siberia (Northern Tungusic) against a much more diversified core region in Manchuria (with Jurchen-Manchu, Udeghe-Oroch, and the Nanai complex). Obviously, the homeland of the extant Tungusic languages must have been located in Manchuria.

The criterion of greatest diversity should, however, not be applied blindly, for the genetic diversity that once existed in the historical core region of a language family can also have been lost for various reasons. A case in point is Turkic, which is linguistically clearly divided into the two principal branches of Common Turkic and Bulghar Turkic. Since Bulghar Turkic (in the form of Chuvash) survives today only on the Volga, the Turkic linguistic homeland could on this basis be mistaken to have been located somewhere in the European steppes. That the Turkic homeland was actually located far in the east, most probably in Mongolia, is confirmed by the presence of Bulghar Turkic loanwords in other eastern languages (especially Mongolic, but also Samoyedic). The model works better for the Common Turkic branch, which, indeed, shows a greater degree of internal diversity exactly in the regions immediately surrounding the presumable eastern homeland.

Some loss of original internal diversity has also taken place in the Sinitic family, for considering only the differences between the modern Sinitic languages, the linguistic homeland would have to have been located very far in the southwest, in the immediate vicinity of the modern Fujian and Guangdong provinces. The available historical information suggests, however, that the expansion centre of Chinese has always been located in the Yellow River basin, from where repeated waves of linguistic influence have spread an entire chronological succession of related speech forms in all directions. Traces of the earliest waves of expansion are, however, only preserved in the southwest, while in the north and west they have disappeared under the later expansion of Mandarin. Even so, the relative internal uniformity of Mandarin suggests that the language is a recent newcomer in many parts of its modern extension, including, in particular, Manchuria.

The role of philological sources is essential when we try to determine the linguistic homelands of Mongolic, Koreanic, and Japonic. The origins of Mongolic can, of course, be connected with the historical Mongols in Eastern Mongolia (the Onon-Argun basins), but the documented former presence of the collaterally related Para-Mongolic languages in Southern Manchuria (the Liao and Shira Muren basins) suggests that the original centre of expansion was located further to the south. By linguistic criteria, this centre of expansion must have been located between the homelands of Turkic (on the west) and Tungusic (on the east). The Bulghar Turkic loanwords in Mongolic were very probably already received before the separation of Para-Mongolic, though the philological verification of this assumption will only be possible when the Para-Mongolic sources can actually be read.

Considering that the earliest linguistically specifiable Turkic (with Bulghar Turkic), Mongolic (with Para-Mongolic), and Tungusic protolanguages were probably spoken at about the same time in adjacent regions in the southern

part of the Mongolian-Manchurian border zone, it is tempting to connect the corresponding protolanguage-level speech communities with actual protohistorical ethnic entities. In such a framework, there are reasons to identify Turkic (especially Bulghar Turkic) with the Xiongnu, Mongolic with the Donghu (later Xianbei), and Tungusic with the Sushen (later Yilou). It is, however, important to realize that the protohistorical ethnic entities mentioned in historical sources are generic references to internally heterogeneous ethnopolitical entities (tribal unions), rather than truly monoethnic and monolingual communities. We will therefore never know how complex the actual linguistic relations within any given protohistorical entity were.

The same is true of the origins of Koreanic and Japonic. While the historical expansion of Koreanic on the Korean Peninsula was connected with the rise of the Silla state (from the southeast), there are strong indications that the neighbouring Baekje state (in the southwest) was predominantly Japonic-speaking until it was linguistically Koreanized. The close connection of the Baekje region with Japan is evident from both archaeological and philological data. Archaeologically, there is no viable alternative to the assumption that the Japonic language family spread to the Japanese Islands together with the Metal Age Yayoi and Kofun cultures (between c. 500 BZ and 500 AZ). Philologically, the former presence of Japonic in Korea is confirmed by the historical Japonic toponyms, documented especially from the Baekje region. Since these toponyms date from a time when Japonic was already expanding on the Japanese Islands, and since they seem to represent a form of speech distinct from the Proto-Japonic lineage, they are best characterized as Para-Japonic. Even so, it is a question of a close relationship.

After locating Mongolic and Tungusic in the southern part of continental Manchuria, and Japonic and Koreanic in the southern part of the Korean Peninsula, it is also possible to approach the question concerning the homelands of Amuric and Ainuic. The cultural vocabulary of Ghilyak shows close connections with Tungusic (especially with Jurchen-Manchu), but in many cases there is no unambiguous way to determine from which of the two language families a given shared cultural word originates. There are also native cultural words in Ghilyak that have no parallels in Tungusic or other languages. This suggests that although Amuric has historically been a marginal and contractive language family, with only a very recent expansion centre in the Tartar Strait region (between the Amur Delta and Sakhalin), it may actually have been relocated to its modern location from Central Manchuria, where it may even have functioned as a local language of 'higher culture' at the time preceding the Tungusic expansion towards the north.

In the Korean context of the Three Kingdoms, the area extending from Central Manchuria southwards was occupied by the Goguryeo state (until 668 AZ) and, later, by its successor state Bohai (698–926). Considering the fact that this was

the territory from where the Jurchen (and Manchu) later emerged, it is possible that Goguryeo and/or Bohai (Barhae) were actually dominated by Tungusic speakers, although there certainly were also other languages. Whether any of these languages were Amuric (or Pre-Proto-Amuric or Para-Amuric) is, of course, impossible to determine, but the possibility cannot be ruled out. Alternatively, Amuric could be connected with the protohistorical entity of Fuyu (Buyo) immediately north of Goguryeo. However this may have been, the protohistorical location of Amuric must, for simple linguistic reasons, have been to the south of its historical position.

Ainuic on the Japanese Islands offers a close parallel to Amuric, for there are reasons to assume that Ainuic has also moved from an earlier more southerly location to its historical location on Hokkaido, Sakhalin, and the Kuril Islands. Although Ainuic (Sakhalin Ainu) is historically adjacent to Amuric (South Sakhalin Ghilyak) on Sakhalin, there is little doubt that Ainuic is originally an insular language family, while Amuric has a continental origin. By all tokens, Ainuic represents a remnant of the languages that were once spoken in Pre-Yayoi (Jomon) Japan. Many of these languages, which may have represented several different language families with a total number of some dozens or even hundreds of idioms, were still present when Japanese started its expansion on Honshu (between 1,500 and 1,000 years ago). In this process, Ainuic was pushed towards the north until it entered Hokkaido, where it replaced the earlier languages of the island. The expansion then continued to Sakhalin and the Kuril Islands. The last stages of this process are to some extent documented archaeologically as well as in folkloric material.

THE ZONES OF INTERACTION

It is not difficult to see that the protohistorical locations of the language families (protolanguages) of Northeast Asia fill a relatively compact and geographically coherent territory. In the core of this territory there are the Liao and Sungari basins of Manchuria, which define the locations of Mongolic on the west (Liaoxi), Tungusic on the east (Liaodong), and Amuric in the north (Sungari). To the south of this core, there is the Korean Peninsula, which was occupied by Koreanic in the southeast (Silla) and Japonic in the southwest (Baekje), with Tungusic probably covering at least parts of the northern half of the peninsula (Goguryeo). On the western and southern margins of the core, in somewhat more diffuse locations, there were Turkic (Mongolia) and Sinitic (the Yellow River basin).

The earlier locations of the language families are confirmed by the traces of mutual interaction between the corresponding protolanguages, as preserved in the modern descendant languages. In general, each language family has had

early contacts with its immediate neighbours: Turkic with Sinitic and Mongolic, Mongolic with Tungusic, Tungusic with Amuric and Japonic, and Japonic with Koreanic. Each of these sets of contact relationship is manifested in one or more layers of early loanwords in the languages concerned. Although in many cases the direction of borrowing is impossible to determine, both linguistic and cultural criteria, when applicable, suggest that the principal direction of lexical flow has been from the inner parts of the continent (Sinitic and Turkic) towards the eastern margin (Mongolic, Tungusic, Amuric), and from continental Manchuria (Mongolic, Tungusic) into the Korean Peninsula (Japonic, Koreanic).

While the other extant language families of the region have clearly a continental origin, Ainuic is an entity that seems to have been located on the Japanese Islands since Neolithic (Jomon) times. This conclusion is, incidentally, coherent with the physical difference between Ainuic and Japonic speakers. A contact between Japonic and Ainuic was only created when Japonic had spread to the part of Japan where Ainuic was originally spoken. This may have been a region (probably in Central Honshu) not too far from the early location of the Japanese state (the Osaka Bay region). When Ainuic was pushed northwards, it ultimately reached Sakhalin and established a contact with Amuric, which, in a similar way, had been pushed northwards along the Sungari and Lower Amur basins.

It has to be noted that the intensity of the early contacts between adjacent language families in Northeast Asia has not been of the same level in every case. For reasons apparently connected with geographical, demographical, cultural, and political factors, some early speech communities interacted more intensively than others. It is also possible that some of the territorial boundaries between the extant families have involved more ethnic and linguistic diversity than can be reconstructed today, a circumstance which may have influenced the intensity of direct interaction. In terms of lexical flow, the most intensive contact zones seem to have been those between Turkic and Mongolic, on the one hand, and between Mongolic and Tungusic, on the other (but not directly between Turkic and Tungusic). It is these relationships of interaction that yielded the corpus of lexical parallels underlying the Altaic Hypothesis. Importantly, the lexical connections of Koreanic and Japonic with the other 'Altaic' languages are much less abundant.

Even when no large corpus of early loanwords is present, however, the areal interaction of the languages of Northeast Asia is visible on the typological level, as manifested in the structural (grammatical) properties of the languages concerned. In a typological classification, the languages of continental East Asia, as a whole, can be divided into two principal types, which may be identified as Altaic and Sinitic, respectively (with both terms used in the typological sense). In the northern part of East Asia, that is, in Northeast Asia, the dominant language

type is Altaic, while most of the languages spoken in the southern part of East Asia, especially in Southeast Asia, are typologically Sinitic. We may also speak of two typological spheres, Altaic and Sinitic, which meet, roughly, in the Yellow River basin.

Without going into a detailed comparison of the actual structural properties of the Altaic and Sinitic language types, we may say that the two types are originally very distinct and suggest two separate centres of typological innovation. Due to the impact of linguistic expansion, both types have spread to new areas, attracting in the process new languages into their spheres. In the meeting zone of the two spheres, we also see mixing of the two types, with features travelling in both directions. Chinese, for instance, is basically a Sinitic language both genetically and typologically, but Mandarin (Northern Chinese) has undergone a process of 'Altaicization', in which it has acquired many features of the Altaic type (such as lexical polysyllabism and incipient suffixal morphology). This process has been long and involves multiple stages of both substratal and superstratal influence, substratal because Chinese has spread northwards to areas where languages of the Altaic type were earlier spoken, and superstratal because Northern China has been conquered several times by northern 'Barbarians' speaking languages of the Altaic type. The last impact of Altaic influence on Chinese was that of the Manchu of the Qing.

Considering the fact that Amuric has been interacting with the languages of Southern and Central Manchuria, which represent Altaic typology, one would expect modern Ghilyak also to be of the same type. As a matter of fact, Ghilyak does have many Altaic typological properties, especially in the morphology and syntax, but it also exhibits surprisingly many non-Altaic features, especially in the phonology and morphophonology. The internal reconstruction of Ghilyak suggests, however, that the non-Altaic features are relatively shallow and have originated only after the earliest contacts with Tungusic. Quite possibly, they are due to the impact of the earlier languages once spoken in the present-day territory of Ghilyak. It is very likely that the earlier languages of the Lower Amur region and Sakhalin represented some other typological complex, distinct from both Altaic and Sinitic.

In the Altaic typological context, Japonic forms a special case. The internal reconstruction of Japonic suggests that the protolanguage had also undergone a process of Altaicization, very much like Mandarin Chinese, but much earlier. There are many indications suggesting that the original typological orientation of Japonic (Pre-Proto-Japonic) was close to the Sinitic type. The language was then Altaicized in the Altaic typological sphere of Northeast Asia, and Japonic entered the Japanese Islands already basically as a language of the Altaic type. The languages of Neolithic Japan, of which only Ainu is historically documented,

represented, however, yet another language type, which, with some caution, may perhaps be labelled Oceanic, or even “Jomonic” in view of the fact that the prolonged coexistence of a variety of languages in Jomon Japan had probably created a distinct local language type. Modern Ainu, in all of its forms, still remains typologically very far from the Altaic sphere, but some of its non-Altaic features are shared by Japonic, which must have got them by interaction with the “Jomonic” languages (not only Ainuic).

The typological prehistory of Japonic is an example of the importance of typology for ethnohistorical conclusions. This is currently one of the most promising fields of ethnohistorical inquiry, since it can potentially yield an understanding of the chronological layering of the structural properties of a language. The layers thus established are not necessarily relevant to linguistic relationships in the genetic sense, but they give indications concerning the previous typological contexts of the language. Typologies, like language families, can be located on the map, and there are indications that typological features are locally more stable than the genetically diagnostic properties of languages, which are easily transferred to new locations. If we know the earlier typological context of a language, we have therefore chances to determine where the language came from.

In the case of Japonic, we see a typological transition from Sinitic to Altaic to “Jomonic” (or Oceanic). The “Jomonic” layer in Japonic was clearly acquired on the Japanese Islands, while the Altaic layer is connected with the former presence of Japonic on the Korean Peninsula. The Sinitic layer, which lies at the bottom, suggests, however, that the Japonic lineage (language family) was even earlier located somewhere within the Sinitic typological sphere, that is, in coastal Continental China. Considering the possible routes from China to Korea, the Shandong Peninsula or the Yangtze River Delta region south of it would appear to be good candidates for the earliest typologically traceable homeland of Japonic.

Assuming that Japonic really entered the Korean Peninsula from Continental China (Shandong) as a language of the Sinitic type, it would initially have been typologically alien in the Northeast Asian context. We do not know whether Koreanic was originally an Altaic type of language family or not, but in any case it certainly had an Altaic typology before Japonic, and it must have been the main local partner for the intrusive Japonic lineage. Most Altaic features of Japonic are therefore likely to be due to the adstratal (possibly also substratal and superstratal) influence of Koreanic. On the other hand, Japonic must also have influenced Koreanic, and there are indeed some features in Koreanic (especially in the prosodic system) which may be due to Japonic (Pre-Proto-Japonic) influence. Finally, substratal Japonic features were also introduced into Koreanic when the Silla language (Old Korean) absorbed the last traces of the Baekje language (Para-Japonic).

From the point of view of absolute dating, it is important to realize that the typological transition of Japonic, and the corresponding geographical movements of the Japonic linguistic lineage, are still relatively recent developments, falling well within the period of Chinese written history. In this context, it is interesting to speculate on the possibility that Sinitic itself may also have undergone a typological transition. Most of the Sino-Tibetan relatives of Chinese, especially in Southeast Asia, are today obvious members of the Sinitic typological sphere, but the languages of the Tibetan branch, and those of several other branches in the Tibetan realm, represent a different typology (which might be called Bodic). It cannot be ruled out that the Sinitic typological sphere was originally formed in the Yellow River basin or south of it before the expansion of Chinese (Proto-Sinitic) to the region. The ancestral form of Chinese may therefore have adopted its Sinitic typology only when it moved into the actual Sinitic typological sphere.

It is still too early to speculate which particular language families may originally have been in the centres of the Altaic and Sinitic typological spheres. It is, however, important to note that the Altaic typological sphere (in this connection also known as Ural-Altaic) also comprises Uralic, a language family which today extends to Northwestern Europe (Fennoscandia and the Baltic region), but which originally seems to have been located in Northeast Asia. The Sinitic type, on the other hand, is consistently represented by the Miao-Yao (Hmongic) and Tai-Kadai (Taic) language families, both of which have historically moved from the central and coastal parts of China towards Southeast Asia. It is possible that these language families already had a Sinitic typology before Chinese. The geographical expansion of the Sinitic sphere in Southeast Asia can be particularly clearly followed in the history of Vietnamese and several other Austro-Asiatic languages. This typological expansion is still continuing today in the border region of China and Southeast Asia.

THE LOST LANGUAGES

What is perhaps most striking about the protohistorical language map of Northeast Asia is that the early languages of which we have knowledge today covered only a small part of the region. Moreover, all these languages were located in the southern part of the region, that is, within the belt extending from Northern China to Southern Manchuria and the Korean Peninsula. This is, of course, no accident, for the situation reflects the fact that the southern part of Northeast Asia comprises several old centres of cultural innovation that have been active since Neolithic times. These centres are typically located in river valleys, notably in those of the Yellow River, the Liao, and the Sungari. The evolution of the linguistic map of the region comprises two steps, which may be summarized as follows:

(1) Expansion: There is no doubt that the Neolithic river valley cultures, relying to a considerable degree on agriculture, involved a previously unparalleled demographic growth in limited areas of production. This growth favoured developments in both internal organization (social stratification and specialization) and external contacts (war and trade), which, in turn, resulted in the adoption of further technological innovations and increased demographic growth. Linguistically, this meant that the speech communities which happened to be in the centre of cultural innovation also underwent rapid growth not only by the factor of internal population growth, but also due to the absorption of neighbouring speech communities by way of linguistic assimilation (language shift). In this way, some speech communities in the region became highly expansive at the expense of others.

(2) Translocation: In Northeast Asia, the primary natural direction of expansion was towards the north, for in the south there were other centres of cultural innovation and demographic growth, especially in the Yangtze River basin, while in the north the population density was lower. The process of expansion was, however, cumulative, so that some of the speech communities located on the northern margins of the river valley cultures also became expansive and started moving towards the north. Pushed by the wave of the primary expansions, these secondary expansions may have triggered tertiary ones, and so on, until the expansions reached physical limits, which in Northeast Asia were formed by the Arctic and Pacific coasts. In this process, the languages previously spoken in the northernmost zone of the region were lost, while the languages that today occupy the northernmost zone have come from the south. All of this need not have involved a significant amount of actual population migration, for basically it was a question of linguistic expansion.

We can actually go even further and speculate that some of the protohistorical language families (protolanguages) located in the southern part of Northeast Asia had themselves been translocated from the south. This possibility is particularly obvious in the case of Japonic, which on typological grounds would seem to have arrived in its protohistorical location on the Korean Peninsula from coastal Continental China, perhaps from the Yangtze River Delta region. The same is actually true of Sinitic, which in its larger Sino-Tibetan context was hardly indigenous to its later location in the Yellow River basin. Considering the overall distribution and typological division of the Sino-Tibetan languages, the best candidate for the ultimate homeland of the language family would seem to be the Sichuan basin, another agricultural zone which, however, lies outside of the Northeast Asian region.

It is tempting to connect the origins of some of the modern language families with actual archaeological cultures. This is often done in the case of Sinitic, which is conventionally linked to the Yangshao Neolithic complex in the 'Central Plains' (5000 to 3000 BZ), often also with the following Longshan tradition

(3000 to 2000 BZ). Similarly, the origins of Mongolic and/or Tungusic can tentatively be derived from the context of the Hongshan Neolithic complex in Southwestern Manchuria (4500 to 2000 BZ). Unfortunately, all such links inevitably remain at the level of unverifiable hypotheses. For Chinese, the only thing we know for sure is that it was the ritual and administrative language of the Shang dynasty (from c. 1400 BZ) in the Middle and Lower Yellow River region. We also know that it had probably come from the south (Sichuan), but we do not know which route it had taken, and which cultural connections it had had on the way.

A general conclusion that cannot be doubted, however, is that all the known expansive language families in Northeast Asia have replaced a variety of older local languages. During the prehistorical period, speech communities must have been generally small and territorially restricted, though they were often mobile depending on the natural conditions. The number of separate speech communities, and quite probably also the number of separate language families, is likely to have reached its maximum in the Upper Palaeolithic. A decline in the linguistic diversity started with the Neolithic period (the Neolithic Revolution), but it took several thousands of years before the change was felt in the most distant corners of Northeast Asia.

In some parts of the region, the loss of languages caused by the expansion of the extant language families must have been massive. In prehistorical Japan, for instance, there may have been up to some hundreds of separate languages, but only one of them, the lineage of Ainuic, has survived until historical times. Similarly, the vast expanses of Siberia today covered by Northern Tungusic (Ewenki-Ewen) must have been the home for a great variety of indigenous languages and language families, of which only one, Yukaghiric (Oduł-Wadul), survives today, translocated far to the northeast. It has to be noted that there are two other language families in Northeastern Siberia (the Bering Sea region), Kamchukotic (Chukchee-Kamchadal) and Eskaleutic (Eskimo-Aleut), but their Eurasian status is ambivalent since they are also linked, both geographically and linguistically, with the New World.

To some extent, the loss of the primary (Palaeolithic) linguistic diversity has been compensated by the secondary diversification of the expansive (Neolithic and later) language families. It is well known that the general language family density of Eurasia is much smaller than that of, for instance, North America, but there is less difference in the corresponding language density rates since some of the extant language families have differentiated into a large number of separate languages. In the Eurasian context, Northeast Asia, with its relatively many isolates and small families, has a special position, even if the diversity in the region does not reach the maximum levels recorded from some other regions (such as New Guinea).

In many cases, the modern diversification of a language family reflects the original linguistic diversity of the region concerned, in that the structural and lexical features of the former (typically mutually unrelated) languages survive in the modern (mutually related) languages in the form of substratal phenomena. The underlying substrata may even be viewed as the cause of the modern diversity. This is even more obvious since we know today that the expansion of a new language is rarely connected with a wholesale replacement of the local population, but involves, in most cases, a process of language shift. It should, however, be stressed that language shift always involves the introduction of a new genetic lineage. No theory of language mixing or creolization can ever blur the distinction between the genetically inherited and the areally transmitted parts of a language.

From the point of view of absolute chronology, it has to be concluded that any reconstruction of the linguistic map of Northeast Asia is inevitably shallow. Recency is the rule of linguistic expansions all over the world, but Northeast Asia is a region that has particularly long remained outside of the sphere of large language families. With considerable reliability, we can follow the expansion of the extant genetic lineages and typological complexes back to protohistorical times, but we can never go beyond the limits of the genetic and typological information. What we do know for sure, however, is that the surviving languages and language families represent but a tiny and randomly selected proportion of the original diversity.

IN LIEU OF A BIBLIOGRAPHY

The present article should be understood as a partial, intermediate summary of the author's views concerning the general mechanisms of linguistic expansion, as observed in the region of Northeast Asia and elsewhere in Eurasia. All too often these mechanisms have been ignored or misunderstood by scholars working on diachronic, comparative, and areal linguistics. Because of the generalizing nature of the text, no bibliographical references will be given. The reader may, however, consult the bibliography in the author's book *Manchuria: An Ethnic History* (Mémoires de la Société Finno-Ougrienne, 222, Helsinki 1996), in which many of the ideas of the present article were first outlined. More recently, the author has dealt with related issues in a number of specialized articles, including "On the chronology of the Ainu ethnic complex" (*Bulletin of the Hokkaido Museum of Northern Peoples*, 11: 1–20, Abashiri 2002), "A framework for the study of Japanese language origins" (*Perspectives on the Origins of the Japanese Language*: 477–490, Kyoto: International Research Center for Japanese Studies 2003), "Tungusic: An endangered language family in Northeast Asia" (*International Journal of the Sociology of Language*, 173: 37–54, Berlin 2005), "The lost languages of Koguryo" (*Journal of Inner and East Asian Studies*, 2(2): 65–86, Seoul 2005), "From Manchuria to Amdo Qinghai: On the ethnic implications of the Tuyuhun migration" (*Tunguso-Sibirica*, 20: 107–119, Wiesbaden 2006), "On typological expansion in the Ural-Altai belt" (*Incontri Linguistici*, 30: 71–83, Pisa 2007), "Some Old World experience on linguistic dating" (*Hot Pursuit of Language in Prehistory*: 223–239, Amsterdam: John Benjamins 2008), "Mongolic as an expansive language family" (*Past and Present Dynamics: The Great Mongolian State*: 127–137, Tokyo: Tokyo University of Foreign Studies 2008), "On the Turkicization of Turkey" (*Finnisch-Ugrische Mitteilungen*, 32–33: 139–150, Hamburg 2009), and "Proto-Uralic: What, where, and when?" (*Mémoires de la Société Finno-Ougrienne*, 258: 57–78, Helsinki 2009).