Contact-induced change in Dolgan : an investigation into the role of linguistic data for the reconstruction of a people's (pre)history

Stapert, E.L.

CHAPTER 1 INTRODUCTION

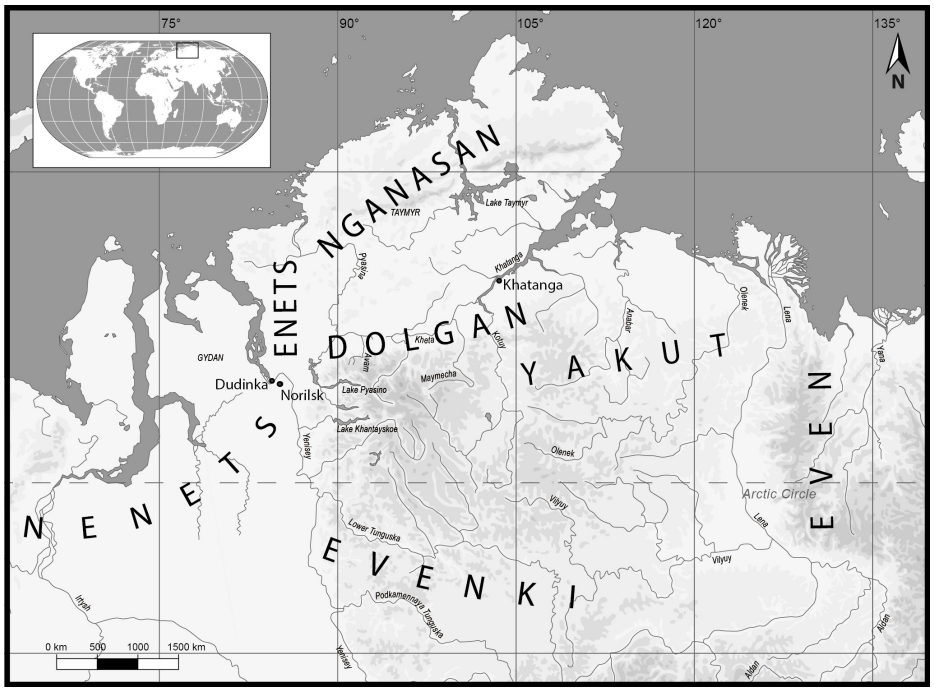
1.1. AIMS OF THE THESIS

Bumping on a wooden sleigh across the endless Siberian tundra, I am beginning to wonder whether it was a sensible idea to start a project in this region with a particular focus on contact. This is the emptiest place I have ever seen. As far as my eye can see from underneath the furry brim of my hat, there is nothing but an endless white desert, devoid of plants, animals or humans and if it had not been for my two Dolgan guides, I would have easily believed that I had landed on a different planet. On days like this, when an incipient snowstorm wipes out the horizon, chasing forth millions of sand-like ice crystals, the white surface below completely merges with the sky above, giving me a feeling of being locked in a giant white egg.

This is the Taimyr Peninsula, the northernmost part of the Eurasian mainland in north-central Siberia. In winter, temperatures can drop below -50 degrees, and snow, ice and strong arctic winds dominate life, or perhaps rather the absence of it, for nine months of the year. For an unaccustomed newcomer to the arctic, like me, it is hard to imagine how humans have been able to survive in a region where access to all elementary needs is extremely restricted. Nonetheless, this area has been inhabited by humans for at least 9,000 years (Denisov 2008: 8),

and regardless of the vast distances and unwelcoming climatic conditions, the history of its inhabitants is characterised by contact.

This applies particularly to one ethnic group called the Dolgan. Not only have they maintained close ties with other ethnic groups throughout their history, but some scholars claim that their very identity as a separate ethnolinguistic group is the 'product' of contact. The Dolgans are one of the six indigenous groups that inhabit the Taimyr Peninsula and the bordering region of the Anabar district (see Map 1) and currently they number 7,885 people (Russian census 2010[1]).



Map 1: Current distribution of indigenous peoples of the Taimyr and neighbouring regions

First mentioned in the 17th century (Dolgikh 1963: 92, Stachowski 1996: 126, Ubryatova 1985: 8), they are the 'youngest' population in this region and unlike their Tungusic, Samoyedic and Turkic neighbours, the Dolgans cannot be unambiguously classified within one of these three ethnolinguistic categories (see Middendorff 1875: 1471, 1476, Castrén 1860, Dolgikh, 1929, 1963: 93, Ubryatova 1985: 5, Anderson 2000: 9, 82); culturally, they share features with the Tungusic

[1] http://www.perepis-2010.ru/results_of_the_census/results-inform.php, accessed on 15-10-2012.

Evenks, and even their ethnonym 'Dolgan' is of Tungusic origin (see Section 2.1 for details). Linguistically, however, they align with their Turkic neighbours, the Sakha (or Yakuts), thus posing a discrepancy between their ethnic and linguistic affiliation. This combination of Tungusic and Turkic characteristics in a single population can only be reasonably explained by the assumption that regular encounters took place between these two groups, and thus through contact.Although interethnic contact is apparent from the current ethnolinguistic profile of the Dolgans, previous accounts differ substantially with respect to ideas about the nature and the extent of this contact, as well as to the moment of appearance of the Dolgans as a separate ethnic group. Some scholars describe them as descendants of groups of Turkic Sakha who adopted a Tungusic life-style of reindeer herding, whereas in other accounts the direction is reversed, and the ancestors of the Dolgans are traced back to Tungusic Evenks who adopted a Turkic language. With respect to the time of their formation as a separate ethnolinguistic group, opinions vary from the early 17th century to as recently as the first half of the 20th century (Dolgikh 1963: 135-139).

The intriguing 'mismatch' between the ethnic and linguistic characteristics of the Dolgans, as well as the conflicting information in the literature about their origins and moment of appearance have been the main incentives to carry out the present study. While the primary focus of this thesis is to approach the history of contact in the Dolgan community from a linguistic point of view, an equally important objective is to interpret the contact-induced changes using historical and ethnographic information as well as insights from language contact theory to infer hypotheses about the most likely social settings in which these changes occurred. In addition, genetic data are employed to underpin the hypotheses about their potential descent with a biological foundation. These genetic data were generated in a project that was undertaken in parallel to the linguistic research, with the explicit purpose to create a context in which linguistic and genetic data could be used to complement each other in inferences about Dolgan ancestry[2]. In short, the main objectives of this thesis are threefold:

1) to identify, describe and analyse contact-induced changes in the Dolgan language.
2) to interpret the linguistic changes in the light of historical, ethnographic and genetic information, as well as insights from language contact theory.

[2] See Whitten et al. (in preparation).

3) to contribute to a better understanding of the role of contact-induced linguistic change in the reconstruction of a people's prehistory, as well as to a better documentation of the Dolgan language.

While contact between Dolgans and Evenks can be traced back as far as the 17th century, and hence is most relevant for the reconstruction of Dolgan (pre)history, this is not the only contact situation the Dolgans have engaged in. A more recent, but extremely influential contact setting has developed since the establishment of the Soviet Union, when Russian influence increased dramatically even in the remotest areas of Siberia. This has had serious repercussions on the social organisation of the Dolgan people and on the use of their native language, as well as on the current situation of widespread Dolgan-Russian bilingualism. Among the younger generations, Russian has become the dominant language, and if no serious measures are taken, the Dolgan language will be replaced entirely by Russian within the next few generations. While contact-induced change as a result of Russian influence is not directly relevant for the reconstruction of Dolgan prehistory, its inclusion in the study is essential in order to build up a complete picture and understanding of the different kinds of contact-induced change in the Dolgan language. In addition, the fact that Dolgan contains the linguistic consequences of two contact situations of a different nature provides a precious opportunity for linguists to scrutinise the proposed correlations between contact situations and their linguistic outcomes within a single community.

For the identification of contact-induced change in Dolgan it is necessary to establish for a certain linguistic phenomenon: a) whether Dolgan has undergone a change; and b) whether the change is contact-induced. In order to do this, Dolgan needs to be compared, on the one hand, with its genealogically closely related neighbour Sakha, also known as Yakut (and other Turkic languages), and on the other hand with its Tungusic and unrelated neighbour Evenki (and other Tungusic languages). Since the specifics of methodology, the nature of the data, as well as the choice of the linguistic material are discussed in Section 1.2 of this introduction, suffice to say here that a phenomenon in Dolgan is considered potentially contact-induced if it is: a) different from Sakha (and the general Turkic pattern); and b) similar to Evenki (and the general Tungusic pattern).

With respect to the second objective, the identified linguistic changes are employed in combination with insights from theories and models of language contact, ethnography, history and genetics to infer information about the likely social setting in which the changes occurred. In language contact theory, there

has been particular emphasis on the establishment of correlations between certain types of contact-induced linguistic changes and the social situation in which they occurred. Based on a diversity of case studies, several models of contact-induced change have been developed over the last sixty years or so, which try to capture regularities in social settings and their linguistic outcomes (e.g. Thomason and Kaufman 1988, Ross 2003, Muysken 2010). A very influential model was proposed by Thomason and Kaufman (1988), who advocate a main distinction between situations of language maintenance and language shift. Language maintenance is a situation where a speech community maintains its native language but 'imports' elements from a contact language, and is typically associated with the copying of linguistic forms (full morphemes). Language shift, on the other hand, is when a community gives up its native language and shifts to another. During this process elements from their native language may be transmitted to the language they are shifting to (the target language), causing changes in the target language (Thomason and Kaufman 1988: 37-50). In contrast to situations of language maintenance, changes associated with language shift are said to have a primarily structural nature. If this correlation is robust enough, the direction of inference could be reversed, and conclusions about the social situation (maintenance or shift) could be inferred from the type of contact-induced change that is found in the language under study: changes in linguistic forms, or substance, would be indicative of a situation of language maintenance, whereas changes in structure would most likely have occurred in a situation of language shift. However, reality shows that these correlations are far from absolute. There are simply too many factors that may influence the linguistic outcome of a contact situation to conceive of such correlations as a relation of cause and effect (see Section 3.2). Thus, while such models can certainly serve as a guideline, careful consideration of the set of contact-induced changes as a whole, as well as inclusion of detailed material from other disciplines, is indispensible.

Two obvious disciplines that fulfil this function are ethnography and history, and their role in this procedure does not require much explanation. After all, it is their main objective to study people and their past, including the relation between different groups of people. However, it will be shown that the information from these disciplines is not always reliable when considered on its own. Like any kind of information, historical and ethnographic accounts may be biased by the aims of the author or by his or her ideological or political background, and therefore must be treated with care. As much as they provide a necessary background for the

interpretation of the linguistic data, they must also be viewed in combination with insights from other domains, so their assumptions can be checked and evaluated.

Probably the most objective kind of information about the past is provided by our genetic material. In our genetic material there are certain parts that remain stable and barely change over time. Therefore, any shared mutations in these parts of our DNA (more specifically the female mtDNA and the male Y-chromosome) are a reliable way to establish common ancestry of individuals, as well as to investigate gene flow between people of different ethnic backgrounds (see Section 2.6 for details). However, while genetic analyses provide specific data on the physical side of the story, including patterns of intermarriage and/or migration, they do not reflect anything about the cultural and linguistic characteristics of the people in question. To conclude, while the data from individual disciplines are informative in their own right, their significance can only be properly evaluated and most importantly increased when viewed from a multidisciplinary perspective. Only a holistic approach will lead to an optimal understanding of the role of the individual elements within the complex mosaic of a people's prehistory.

1.2. DATA COLLECTION AND METHODOLOGY

1.2.1 CHOICE OF FIELD SITES AND DATA COLLECTION

The linguistic data for Dolgan were collected during three fieldtrips to the Taimyr Peninsula in the villages Volochanka (2008), Kheta (2009, 2010), Syndassko (2009) as well as in the towns of Khatanga (2009) and Dudinka (2010). The first trip took place from June until September in 2008, but because of the opaque procedures of Russian bureaucracy I could spend only the final month on the Taimyr Peninsula. I spent this time in the village of Volochanka, and due to the restricted time and because I had no reason to be confident that I would make it through the bureaucratic maze again, I devoted most work to the collection of as much Dolgan language material as I could possibly get. The data included narratives, the Pear Stories (see Section 1.2.2), and some grammatical elicitation. The preceding two months of this trip were spent in the village of Baajaga (Taatta District, Sakha Republic) where I collected additional material for Sakha, in particular the Pear Stories for comparison with Dolgan.

The second trip, which took place from February to May in 2009, went more smoothly from an administrative point of view, and this time I was able to exhaust my fieldwork time, and I spent the full three months in the villages of Kheta and Syndassko, and the town of Khatanga. During this trip, I collected the core part of the database of Dolgan narratives, and completed most of the transcriptions and translations of the recordings. In addition, I carried out elicitation on grammatical topics with the help of questionnaires (see Section 1.2.2).

The third and final trip took place in the summer of 2010 (July - August) and was primarily intended to collect detailed information on the Dolgan lexicon. The second goal of this trip was to fill in gaps in the data already collected, to cross-check transcriptions and translations, and to eliminate any open questions in the database.

The choice of fieldwork locations was guided by the motivation to collect language material from different dialects of Dolgan as well as from speakers who differ in their linguistic dominance and their level of bilingualism in Dolgan and Russian. The Dolgan language can be divided into two dialects - the upriver dialect, spoken in upper region of the Kheta River in the southwestern region of the Taimyr Peninsula, and the downriver dialect, spoken towards the northeast down the river Kheta and in the Khatanga basin (see Map 2 in Section 1.3.1). The geographical distribution of these dialects suggests that the western (upriver) dialect may have undergone more influence from Evenki since it is currently closer to an Evenki-speaking area, whereas the eastern (downriver) dialect may have retained more similarity to Sakha. While the decision to visit a western, an eastern and a middle village was my own conscious choice, the actual villages I ended up in were rather determined for me by the people who happened to provide transport. Since helicopters may go only every two or three weeks, and trucks do not drive when it is colder than -40 degrees, one cannot be too selective in this respect. Thus, the narratives recorded in Volochanka represent the upriver dialect, the ones from Syndassko the downriver dialect, and the recordings from Kheta the dialect of the transitional region.

1.2.2. THE DATA

The main body of the data consists of narratives, produced by native speakers of Dolgan. For the collection of these data, speakers were asked to tell a story about a

topic of their choice, and only if they had trouble coming up with something would I make suggestions of potential themes. The stories were recorded with an external microphone (AKG C 1000 S) and a digital Marantz recorder (PMD660) in PCM format with a sample rate of 48 kHz and a sample size of 16 bits.

They were further processed using the transcription and interlinearisation software ELAN[3] (Sloetjes and Wittenburg 2008) and Toolbox (SIL international). Segmentation of the soundfiles, for which the intonational sentence was taken as a segmentation unit, was done in ELAN. The audio data were transcribed using a Latin-based transcription system instead of the official Cyrillic-based Dolgan orthography. There are several motivations for this choice. First, the Latin-based system, developed by Pakendorf for the transcription of Sakha texts (Pakendorf 2007), allows for a better representation of phonetic variation in the oral texts. Since an important value of spoken texts is to capture this variation, this Latin-based system was a better choice than the Cyrillic orthography. Second, the Latin-based system corresponds to the transcription system used in the database for Sakha compiled by Brigitte Pakendorf (see Pakendorf 2007 for details). Since an important component of the present research involves comparison of spoken texts of Dolgan and Sakha, the use of an identical transcription system facilitates this task considerably. After transcription, the texts were translated into Russian, and interlinearised using Toolbox, applying where possible the glossing system prescribed by the Leipzig Glossing Rules[4].

The transcription and translation into Russian were done under the watchful eye of native Dolgan speakers. In fact, they did most of this work and my part was to understand their explanations and enter the data into the computer. It is obvious that the current corpus would not exist without their invaluable help and patience. Most texts were double-checked with a second speaker to verify the translations. Interlinearisation in Toolbox was done by me, but not without frequent consultations with Dolgan speakers in cases of uncertainties and ambiguities. The only phase of processing in which the Dolgan people were not involved was the additional translation into English, to make the texts accessible to a larger community. The three field trips have resulted in a database of Dolgan of over 3 hours of narratives, containing 16,250 words. It comprises 22 stories, narrated by 15 different speakers of both sexes, ranging in age from 8 to 76.

---

[3] http://tla.mpi.nl/tools/tla-tools/elan/ developed at the Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen.

[4] http://www.eva.mpg.de/lingua/resources/glossing-rules.php.

In addition to spontaneous narratives, I collected a number of the so-called Pear Stories (Chafe 1980) in Dolgan, as well as in Sakha. These are semi-spontaneous narratives, prompted by a short silent film that I showed on my laptop[5]. After the screening of the film, the participant is supposed to retell the events that he observed in the film. The rationale behind this method is that it should allow the linguist to collect linguistic material that is relatively comparable across individuals and languages in terms of theme, vocabulary and narration structure, without relying solely on data from questionnaires. For Dolgan 7 Pear Stories were collected with a total number of 1,427 words and for Sakha 9 of these semi-spontaneous narratives were recorded with a total number of 1,840 words. While this procedure yields results that are certainly better than translated sentences, it works only to a certain extent. First, people vary considerably in their understanding of the task, as well as in the interpretation of the filmed events, which can still result in very different stories despite the identical stimulus. Another problem of the Pear Story is that it is culturally quite specific and therefore the level of 'naturalness' in retelling this story may vary across geographical regions and cultural settings. For example, the prominently figuring pears are not unmarked objects in the arctic, and they caused initial confusion, not in the least because there is no native Dolgan word to describe them. Also the goat was typically identified as a reindeer due to the lack of goats in the arctic region. Nonetheless, Pear Stories are valuable material for certain purposes, in particular for frequency counts of certain forms or constructions, because it is the closest one can get to a collection of comparable narratives.

In addition to the spontaneous and semi-spontaneous narratives, which constitute the largest part of the database, elicitation tasks were conducted in order to investigate certain linguistic domains in detail, as well as to provoke explicit statements on the (un)acceptability of certain linguistic forms and constructions. They proved particularly necessary in the study of relative clauses,

[5] This film was developed by Wallace Chafe in 1975, and his own results were published in his 1980 book. In short, the film starts with a scene in which a man is picking pears (hence the Pear Story). First a boy with a goat comes by, followed by a boy on a bike who steals one basket of pears. As he cycles off with the pears, a girl on a bike approaches him from the opposite direction and in passing snatches his hat. The boy is distracted, hits a stone and falls from his bike, the pears rolling over the road. Three boys arrive who help him collect the pears and they return his hat. In return they receive a pear each. In the final scene, the pear farmer is shown, descending his ladder. He discovers that one basket of pears is missing, and at the same moment three boys pass by, each munching a pear. He stares after them looking confused.

since these do not occur frequently in spontaneous speech in Dolgan, as is the case in many languages. In the examples adduced in this thesis, it is always indicated when elicited material is used. Like the narratives and the Pear Stories, the elicited material was interlinearised, added to the digital database and sorted by elicitation topic.

The final kind of collected material concerns the lexicon. While narratives provide a rich source of information about lexical items and their semantics, they are not sufficient for the study of the lexicon at a more specific level. If one is interested in rare lexical items or in fine-grained semantic differences between lexemes, explicit elicitation is the only way to access this information. Since not only the lexical forms, but also their semantics may change in a situation of contact, an in-depth study of part of the lexicon was carried out to investigate how this domain was affected in Dolgan.

For this purpose a wordlist was used that was originally designed for the Loanword Typology project (Haspelmath & Tadmor 2009). Importantly, this wordlist had already been elicited for Sakha (Pakendorf & Novgorodov 2009), therefore elicitation in Dolgan allowed for a direct comparison of these items in both languages. The Loanword Typology list was initially designed to investigate 'borrowability' of the included meanings in a sample of 41 languages. The 1,500[6] meanings are distributed over 24 semantic fields, ranging from non-cultural items, such as body parts, to highly culturally-determined lexicon, such as technical and educational concepts. The Loanword Typology list itself is based on the International Dictionary Series (an ongoing project founded by Mary Ritchie Key (1924-2003) and now headed by Bernard Comrie) and the Swadesh 207 list, both of which are intended to compare lexicon across languages. The entire list of items was elicited with one Dolgan speaker in Dudinka. However, whenever she was not entirely sure about a form or meaning, she did not hesitate to use her network of Dolgans in town and consult other speakers. Certain parts were double-checked later with speakers in Kheta. The meanings, implications and semantic nuances were discussed at great length before they were entered into a searchable Filemaker database.

Additional data on the Dolgan language that were not collected by me personally are extracted from published grammars and dictionaries. The

[6] This is the latest version of the Loanword Typology list. The version used in the comparative handbook 'Loanwords in the world's languages' (Haspelmath & Tadmor, 2009) consists of 1,460 items.

grammars are 'The language of the Norilsk Dolgans'[7] (Ubryatova 1985) and 'The Dolgan language'[8] (Artemyev 2001). The used dictionaries are 'Dolgan Lexicon'[9] (Stachowski 1993), 'Dolgan Lexicon, supplementary volume'[10] (Stachowski 1998) and the dictionary by Aksenova, Beltyukova and Kosheverova (1992).

The data on Sakha are primarily taken from texts from the spoken corpus that was compiled by Brigitte Pakendorf between 2002 and 2006, and that I was kindly granted access to. This corpus contains 5 hours of annotated texts of mainly life stories (29,400 words), which were recorded from 15 speakers from different regions of the Sakha Republic, representing a variety of Sakha dialects[11]. Only the Sakha Pear Stories were collected and processed by me. Additional data were taken from grammars (Kharitonov 1947, 1960, Korkina 1970, Pekarski [1907-1930] 1958-1959, Ubryatova 1982).

In the absence of an accessible spoken corpus for Evenki, for this side of the comparison I had to rely on published sources. The same is true for the comparison with other Turkic and Tungusic languages. For Evenki, the main sources were the Evenki grammars by Nedjalkov (1997) and Bulatova and Grenoble (1999) and dictionaries by Vasilievich (1968), Boldyrev (1994), Myreeva (2004); for the analysis of texts the collections of folkore texts by Vasilievich (1936, 1966) were used. While I am clearly aware that written sources alone are not ideal for detailed study of grammar and lexicon, and that the results may be improved by more targeted fieldwork, particularly with respect to semantic details of lexical items in Evenki, they were the best available resource to complement my own data.

1.2.3 METHODOLOGY FOR IDENTIFYING CONTACT-INDUCED CHANGE

Although the applied methodology varies slightly for each phenomenon under study, this section serves to elucidate the general principles that are applied

[7] язык норильских долган.

[8] долганский язык.

[9] Dolganischer Wortschatz.

[10] Dolganischer Wortschatz, Supplementband.

[11] The texts in the corpus were collected in the districts of Olenek, Verkhoyansk, Suntar and Taatta.

throughout the thesis to: a) identify a linguistic change; and b) to determine if it is contact-induced.

As will be discussed in more detail in Chapter 3 (Section 3.2) it is almost never possible to give solid proof that a change is contact-induced, unless clear lexical copies are concerned. For many changes in phonological or morphosyntactic structure a language-internal motivation cannot be excluded, even in cases where contact seems a very plausible explanation. However, according to Thomason this is not a problem, since the two kinds of explanation need not be mutually exclusive. Instead, Thomason and Kaufman (1988: 57, 61), and with increasing emphasis Thomason (2001: 62-63 and 2010), stress the importance of multiple causation in the explanation of language change, including both internal and external motivations:

> The best explanation for any linguistic change will take all discoverable causal factors into account, both internal and external. The rather extensive literature that attempts to decide between an internal and an external cause of a particular change is a waste of effort - the dichotomy is false, and the best historical explanation might well have to appeal to both causes. (Thomason 2010: 34)

Despite the fact that we may not be able to give absolute proof of contact-induced change, and conclusions have to remain tentative, it is still possible and mandatory to set up the necessary (but not sufficient) conditions that should be met if an external explanation is considered. According to Thomason (2001: 93-94, 2010: 34) these are:

1) identify the recipient language and consider the entire system of this language rather than individual subsystems.
2) identify a source language
3) find shared features between the source language and the recipient language
4) prove that these features are old, and are not an innovation, in the source language
5) prove that these features are new in the recipient language, and were not present before it came into contact with the source language.

If only a subset of these conditions is met, an explanation can at best be tentative. Formulated in a slightly different fashion, and putting less emphasis on Thomason

and Kaufman's advice to consider the language system as a whole, Heine and Kuteva (2005: 33) come to a similar conclusion when they define contact-induced change as follows:

> If there is a linguistic property x shared by two languages M [Model Language, E.S.] and R [Replica Language, E.S.], and these languages are immediate neighbors and/or are known to have been in contact with each other for an extended period of time, and x is also found in languages genetically related to M but not in languages genetically related to R, then we hypothesize that this is an instance of contact-induced transfer, more specifically, that x had been transferred from M to R. (Heine and Kuteva 2005:33)

In the present study, the same methodological principles are taken as a guideline for the identification of contact-induced change in Dolgan. However, due to the linguistic situation on the Taimyr, as well as the character of the available data, the order of procedure has been adapted.

The first step for the identification of contact-induced changes in Dolgan has been the establishment of differences between Dolgan and its genealogically most closely related neighbour Sakha. This goal was achieved through a careful comparison of the two languages on the basis of the spoken text corpora mentioned in Section 1.2.2. The diagnosed points of divergence between Dolgan and Sakha were then compared with other Turkic languages in order to clarify which of the two languages behaves 'typically' for the family and which one does not. In cases where Dolgan turned out to be the deviant language, the relevant construction was compared with its functional equivalents in the other, unrelated, neighbouring languages including Tungusic Evenki and Samoyedic Nganasan. In practice, comparison with Evenki proved to be most relevant because we know from history that these two ethnic groups have been in contact for a long time, and that this interaction has been important for the formation of the Dolgan people (see Chapter 2). In a similar fashion to what was done for the Turkic language family, comparison of Evenki with other Tungusic languages was carried out to evaluate whether the structures in Evenki are typical for the Tungusic family and to exclude the possibility that Evenki has undergone contact-induced change. Following the reasoning expressed in the quote by Heine and Kuteva, the idea is that if the pattern in Dolgan differs from genetically related Sakha, but matches the pattern in genetically unrelated Evenki, then there is good reason to

consider transfer of the phenomenon under study from Evenki to Dolgan as a potential explanation of this difference.

1.2.4 CAVEATS

While the procedure described above represents the ideal scenario, in practice a number of caveats are in place. First, not for every phenomenon that looks as if it might be contact-induced do we have the full range of comparative material available, due to the lack of detailed description. In these cases, conclusions must remain speculative, and only more in-depth work on the other relevant languages may be able to eliminate this uncertainty in the future.

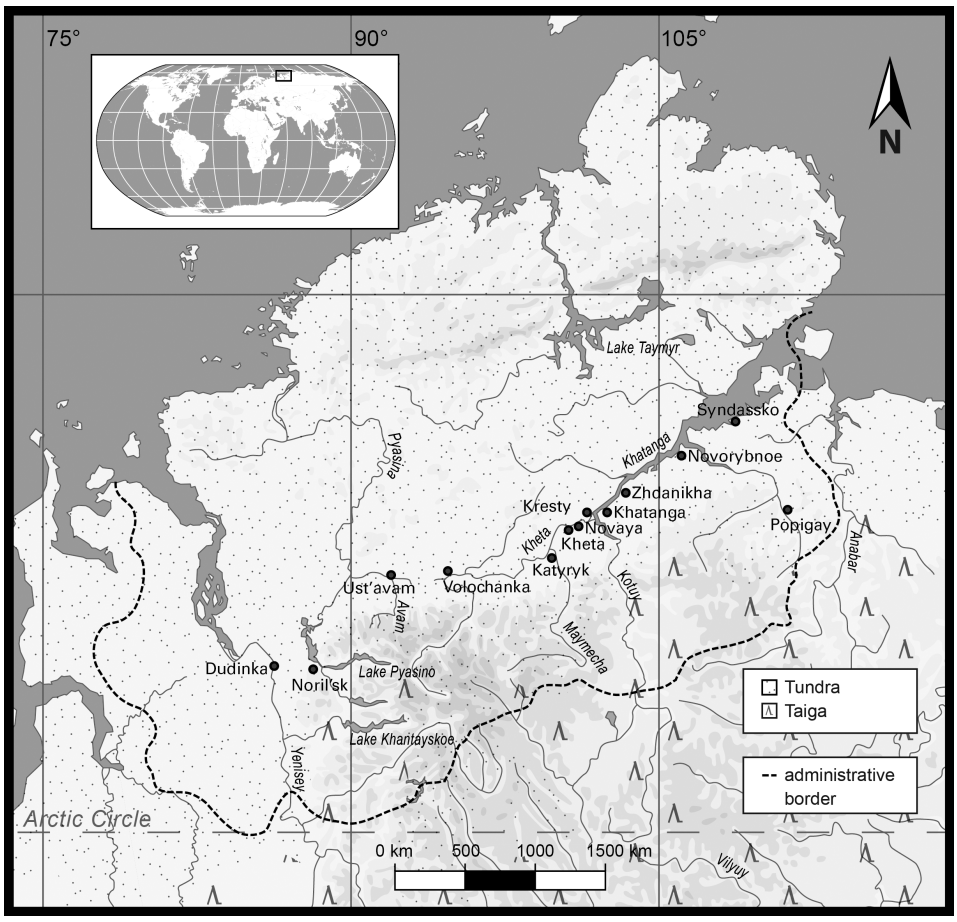
Second, there are differences in the nature of the data under comparison. While spoken corpora were available for Dolgan and Sakha, this was not the case for the languages they were compared with. To my knowledge, there is no accessible corpus of spoken annotated Evenki or Nganasan texts, which would have been necessary for a homogeneous dataset. One possible way to circumvent this problem would have been to use written data for Dolgan and Sakha as well. However, while evading a problem on the one side, another, more serious problem would have been incurred on the other: the grammar of written Dolgan is heavily subjected to the standards of literary Sakha, and since the differences between Dolgan and Sakha are the focus of my interest, research based on written data would thus have made the entire enterprise pointless. As a second best, the lack of spoken data was compensated by written Evenki texts, as well as two published descriptive grammars (Nedjalkov 1997 and Bulatova and Grenoble 1999). The same applies to comparative data for other Turkic and Tungusic languages. For the Southern Tungusic language Udighe a number of transcribed and annotated spoken texts were available in the Udighe grammar (Nikolaeva & Tolskaya 2001).

1.3. DESCRIPTION OF FIELD SITE

1.3.1 GENERAL INFORMATION

The largest proportion of the fieldwork for this study was conducted in the villages of Volochanka, Kheta and Syndassko, three villages situated on the Kheta River. In addition, some work was done in the towns of Khatanga, Dudinka as well

as the village of Levinskie Peski, which is across the Yenisey River from Dudinka (see Map 2). Volochanka, Dudinka and Levinskie Peski belong to the administrative district of Dudinka, whereas Khatanga, Kheta and Syndassko belong to the district of Khatanga.



Map 2: Dolgan villages and towns on the Taimyr Peninsula (2012)

All villages are situated on the high riverbanks of the Yenisey and Kheta Rivers. The latter changes its name into the river Khatanga as it flows north. This means that they can all be reached by boat, but the long journey, changing water levels and the short period of time in which the rivers are free of ice makes this means of transport unattractive. Instead, Volochanka and Syndassko are typically reached by helicopter, but services are often dependent on an unpredictable

interplay of obscure factors. In theory, flights to these places take place every two weeks, but due to bad weather, high fuel prices, (alleged) lack of passengers and mood fluctuations of the people in charge this schedule can best be described as a good intention. As a result, many locals use private means of transport, which are typically snow scooters or extraterrestrial-looking all-terrain vehicles. In the tundra, reindeer or dog sleds are also still used. While the advantage of private transport is that it gives more freedom as to the moment of travel, the disadvantages are that a journey by land to the closest town involves three days instead of two or three hours, and it is a risky enterprise. Due to the absence of roads in this area, the extreme weather conditions, and the state of the vehicles, accidents occur regularly, and in the worst case a traveller never reaches his destination. In contrast to Volochanka and Syndassko, Kheta is not served by helicopters and is frequented by boats in summer and by taxis in winter. In this season, the frozen river is used as a road. While the availability of transport is an issue for this village as well, its relative closeness to the town of Khatanga makes access a bit easier.

The size of the villages varies from about 400 to 600 people, and the majority of the inhabitants are Dolgans. Only in Volochanka is there an almost equal number of Nganasan people. A summary of the ethnic composition of the villages, as well as of the larger towns Dudinka and Khatanga, is given in Table 1.1[12].

Table 1.1. Ethnic composition of fieldwork locations

	Dudinka	Lev.Pesk.	Voloch.	Khatanga	Kheta	Syndassko
Dolgans	1.715	97	295	788	362	513
Nganasan	654	7	266	13	4	3
Evenks	260	0	0	2	0	5
Nenets	550	7	3	5	2	0
Other	20.855	59	40	2.126	16	5
Total	24.034	170	604	2.934	384	526

The current set-up of the villages is of a quite recent make. Traditionally, the Dolgans led a nomadic lifestyle and lived in tents made out of reindeer hides, or in *baloks* - little huts on sleigh runners, covered with tarpaulins, which were moved from camp to camp by harnessing reindeer to the front of them. Only after the

[12] These numbers are taken from the data provided by the official website of the Taimyr www.taimyr24.ru, accessed on September 26, 2012.

forced settlement by the Soviet Regime, which took place in the 1970's, did the villages get their current shape. They consist of long one-storey, barrack-like houses built parallel to the river, which typically contain four apartments, providing room for four families. Every village in which I stayed had a hospital, a school, a library, a post office, a club house, and a diesel station for generating electricity. Volochanka and Syndassko had a state-owned shop, but Kheta only had two cabins in which private merchants resold goods from Khatanga at astronomic prices. Most people survive exclusively on the reindeer meat and fish that they either catch themselves, or get from family or friends.

1.3.2. SOCIOLINGUISTIC SETTING IN THE VILLAGES

The villages I visited differ considerably with respect to the use of the Dolgan and the Russian language, as well as the attitudes towards the use of each language. The information provided here will be repeated in the relevant places for the sake of argumentation, but since the level of bilingualism, differences in linguistic dominance and speaker attitude are important factors for the study of contact-induced change, it is worth summarising this information in one place for quick reference.

As mentioned before, all Dolgan speakers are bilingual in Russian. Nonetheless, I observed a difference in the linguistic, as well as cultural, dominance of Russian across the villages. The shorter the distance to the towns, the more the influence of Russian language and culture has made itself felt and heard. Thus, of the three villages in which I recorded most of my material, the speakers in Syndassko have retained the highest percentage of Dolgan-dominant speakers, Volochanka the lowest, and Kheta occupies an intermediate position between the two, thus matching their geographical distribution in relation to the town of Dudinka. The sociolinguistic situation in these villages is briefly discussed below.

Wherever I expressed my wish to learn about the Dolgans and their language, people unanimously advised me to go to the village of Syndassko on the border with Yakutia, where in their opinion Dolgan language and culture are best maintained. Although everybody in Syndassko is bilingual in Russian, Dolgan is still widely used on a daily basis and is vital for communication across all generations. While children are exposed to Russian from a very young age through

television and through the omnipresent bilingualism in the community, there are still children who grow up predominantly monolingual in Dolgan for at least the first five years of their life. These are mostly children of semi-nomadic reindeer herders, who migrate with their nuclear family around the tundra, and whose dominant language has often remained Dolgan. Apart from a certain amount of Russian that these children hear when they visit the village, they will have their first serious encounter with Russian only when they enter school, where Russian is the language of instruction. Children who grow up in the village also learn Dolgan as a first language, but they will have had more exposure to Russian before they reach school age, through television and through organised events in the village, for which Russian is also regularly used. Therefore their 'bilingual life' starts slightly earlier than with the semi-nomadic children. Even after entering the school system, Dolgan remains the main language of communication in many domains within, as well as across generations in Syndassko. The considerable amount of code-switching between Russian and Dolgan for particular topics does not seem to greatly affect people's proficiency in Dolgan, and apart from the use of some Russian lexical items, they seem to keep the languages apart rather well.

With respect to language attitude, the Dolgans in Syndassko seem rather proud of their native tongue. While they consider knowledge of Russian necessary for practical purposes, and especially for education, everyone I spoke to defined Dolgan as their native language and spoke of it in a positive way. It is obviously no coincidence that this linguistic situation obtains in Syndassko, which is the most remote village, geographically and culturally, with respect to contact with monolingual Russian-speaking communities. Its remoteness from Russian, in combination with its geographical and cultural proximity to the Sakha Republic, where the position of the closely related Sakha language is much stronger, may explain the unique preservation of traditional habits and language in this area.

A very different situation applies in Volochanka. In this village, Dolgan is irreversibly in decline and will most probably disappear within the next couple of generations. At present, fluent Dolgan speakers are hard to find, and the few exceptions are restricted to the generation that is now over 60. People between 40 and 60 master the language to different degrees, but all are beyond doubt dominant in Russian. The current situation in this village may be influenced by its location and the concurrent socio-historical developments that took place. Its vicinity to Dudinka and a slightly more friendly landscape than the naked tundra around Syndassko may have resulted in a stronger Russian presence in this area

from quite early on. During Soviet times, Volochanka was the administrative center of the area, and there was a large state farm where the indigenous population encountered many Russians, and where they worked in a mixed ethnic community, in which Russian was used as the means of interethnic communication, or the lingua franca. This high activation level of Russian may have been maintained because Volochanka is inhabited by both Dolgans and Nganasans (see Table 1.1), who communicate with each other in Russian. In Volochanka, hardly any child speaks Dolgan, and while their parents observe this with a certain melancholy, no action is undertaken to change this tendency. In contrast to Syndassko, people in Volochanka commented that it is a pity that the children do not learn the language anymore, but continued that it would be of no use anyway. If they learn Russian well, so most people said, they will be able to study well and perhaps have a better life.

The third location, Kheta, could be seen as the middle ground between the two extremes to its east (Syndassko) and its west (Volochanka). In Kheta, Dolgan is still spoken well by the older generation, and for some of the oldest people, Dolgan may still be the dominant language. However, they are very few. In addition, their attitude towards the Dolgan language is more positive than in Volochanka, and the teachers in the school are devoted promoters of the Dolgan language and culture. In a similar fashion to Volochanka, increasing age typically correlates with increasing proficiency in Dolgan for the age group between 40 and 60. The older people speak it better and more frequently than the younger ones. Typically, children do not learn the language from their parents anymore, but they have a reasonable passive understanding, and actively use standard expressions like *kel* 'come here' or *bar* 'go away'. In a few exceptional cases parents do speak Dolgan with their children, as for example did my main consultant, but this is not characteristic of the situation in the village as a whole. Russian is rapidly becoming the dominant language, and is undoubtedly already so in the age group under forty. Everybody in the village has (near)-native command of Russian and people describe themselves as 'Russified Dolgans'.

1.3.3 THE CONSULTANTS

During the compilation of the corpus, the intention was to collect narratives from Dolgan speakers across a wide range of geographical locations, age, gender and

language proficiency. As a result, the corpus includes speakers from four geographical locations (Volochanka, Kheta, Syndassko and Dudinka), ranging in age from 8 to 76, and including both males and females. All speakers are native speakers of Dolgan, and are bilingual in Russian, since nowadays it is virtually impossible to find monolingual speakers of Dolgan. Only one elderly woman in the village of Syndassko sometimes had trouble expressing herself in Russian, but still knew it well enough for ordinary conversation. Most speakers master both languages very well, but in areas where Russian is prominently present, the percentage of Russian lexical items in the Dolgan speech of the consultants is higher than in areas where this influence is limited. While both men and women are included in the sample of speakers, the predominance of elderly women in linguistic activity was inevitable in this particular fieldwork setting. An overview of the speakers, who are referred to by their initials for reasons of confidentiality, is given in Table 1.2[13].

Table 1.2. Overview of Age, gender and location of consultants

Location of recording	Initials	Age	Gender
Volochanka	LKS	63	F
	EIB	74	F
	IVA	55	F
	ANS	45?	M
Kheta	TJP	40	F
	MIC	75	F
	SNB	71	F
	APF	70	M
Syndassko	IMA	47	F
	PPK	74	F
	DPK	9	F
	APC	63	F
	MSA	8	M
	SEK	14	M
	NMC	53	M
	SSK	19	M
Dudinka/	ESB	76	F
Levinskie Peski	LSB	59	F
	TIS	52	F

[13] This list includes consultants who provided narratives as well as Pear Stories, therefore the number of speakers is higher than the number of 15 mentioned before, which included narratives only.

1.4. OUTLINE OF THE THESIS

One of my strongest opinions regarding the study of contact-induced change is that it can only be properly understood if it is embedded in the broad context of the social history of the communities in question. Therefore, Chapter 2 provides a detailed picture of the geographical, historical, ethnographic and linguistic characteristics of the Dolgan people and their ancestors. In addition, it gives an overview of the main results from the analysis of Dolgan DNA-samples, which complements this linguistic analysis.

Chapter 3 gives an overview of the field of contact linguistics and introduces essential concepts used in language contact theories and in the study of contact-induced change in general. Rather than trying to cover all of the different theoretical frameworks that have been proposed, I chose to elaborate a selection of ideas that have proved relevant and insightful for the analysis of the Dolgan data. The chapter concludes with a discussion of the role of language contact theory in the study of contact-induced change.

Chapter 4 investigates lexical change in Dolgan. After an introduction to the analytical framework that is employed for the definition and analysis of lexical change, in which six types of lexical differences between Dolgan and Sakha are identified, the types of difference are analysed in both a quantitative and a qualitative way. For the quantitative analysis, first the proportion of differences between Dolgan and Sakha is determined for 24 semantic fields to determine whether the differences and potential contact influence are concentrated in certain semantic domains. After that, the focus shifts towards the analysis of the six types of difference themselves. The relative frequency of the different types is investigated and it appears that the most common type of difference between Dolgan and Sakha is semantic change. Therefore, this type is then investigated in detail, uncovering important changes in the semantic structure of kinship terminology as well as the semantic field of 'the body' that most probably developed as a result of contact with Evenki. The second type of difference that is zoomed into is replacement, analysing copies from both Evenki and Russian.

Chapter 5 discusses differences in the inflectional paradigms of the auxiliary verb *e-* 'to be' and of unstable noun stems. These paradigms show irregular inflection in Sakha, whereas in Dolgan they have developed a regularised alternative. While explicitly leaving room for a language-internal explanation, it is

argued that this regularisation may have been accelerated by Evenks who learned Dolgan as a second language.

In Chapter 6 the habitual participle is examined. Analysis of the morphosyntactic properties of this participle, as well as of its frequency of use, shows that Dolgan and Sakha differ significantly in both respects. In contrast to Sakha, where the participle is used with a verbal as well as with a nominal function, the nominal use in Dolgan does not occur. However, the verbal use of the participle occurs with a much higher frequency than in Sakha. Although more research is needed to confirm this hypothesis, it is noted that the use of the habitual participle in Dolgan is more similar to the morphosyntactic properties of the habitual in Tungusic languages than its use in Sakha.

Chapter 7 discusses word order patterns, showing that Dolgan allows much more flexibility in this domain than Sakha. Instead of applying strict SOV order as do most Turkic and Tungusic languages, the spoken Dolgan text corpus reveals a high percentage of SVO order. While a language-internal explanation for this difference cannot be excluded, a more plausible explanation seems to be the increasing dominance of Russian, in which SVO is the unmarked word order.

Finally, in Chapter 8 differences in clause combining strategies are analysed. These appear to be rather diverse, and it is argued that some of them could be the result of contact with Evenki, whereas the majority seems to have developed more recently as a result of the increasing linguistic dominance of Russian, as well as language attrition. Due to the complex combination of relevant social factors and the diversity of linguistic outcomes this chapter, in particular, highlights the importance of multi-causality in the explanation of contact-induced change.

Chapter 9 offers a detailed discussion of the conclusions reached in the individual chapters, embedding the linguistic results in the historical, ethnographic and genetic context presented in Chapter 2, and viewing the set of changes as a whole. By taking this holistic view I work towards a synthesis of these different disciplines to build up a more complete picture of the prehistory of the Dolgans.

Chapter 10 concludes the thesis with a brief conclusion and an outlook for future research.