



Universiteit
Leiden
The Netherlands

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.

Citation

Stapert, E. L. (2013, September 26). *Contact-induced change in Dolgan : an investigation into the role of linguistic data for the reconstruction of a people's (pre)history. LOT dissertation series*. Retrieved from <https://hdl.handle.net/1887/21798>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/21798>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/21798> holds various files of this Leiden University dissertation.

Author: Stapert, Eugénie

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

Issue Date: 2013-09-26

CONTACT-INDUCED CHANGE IN DOLGAN:
AN INVESTIGATION INTO THE ROLE OF LINGUISTIC DATA FOR THE RECONSTRUCTION
OF A PEOPLE'S (PRE)HISTORY

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6006
e-mail: lot@uu.nl
<http://www.lotschool.nl>

Cover illustration: Portrait of a Dolgan child. Photo by Eugénie Stapert

ISBN: 978-94-6093-119-2
NUR 616

Copyright © 2013; Eugénie Stapert. All rights reserved.

CONTACT-INDUCED CHANGE IN DOLGAN:
AN INVESTIGATION INTO THE ROLE OF LINGUISTIC DATA FOR THE RECONSTRUCTION
OF A PEOPLE'S (PRE)HISTORY

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus Prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 26 september 2013
klokke 16.15 uur

door

Eugénie Stapert
geboren te 's-Gravenhage, Nederland
in 1981

Promotiecommissie:

Promotores: Prof. dr. M.P.G.M. Mous
Dr. dr. B. Pakendorf, Directeur de Recherche (C. N. R. S Lyon)

Overige leden: Prof. dr. É. Á. Csató Johanson (Uppsala Universitet)
Prof. dr. A. Backus (Tilburg University)
Dr. M. G. Kossmann

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	I
REFERENCE INFORMATION	IV
CHAPTER 1 INTRODUCTION	1
1.1 AIMS OF THE THESIS	1
1.2 DATA COLLECTION AND METHODOLOGY	6
1.2.1 CHOICE OF FIELD SITES AND DATA COLLECTION	6
1.2.2 THE DATA	7
1.2.3 METHODOLOGY FOR IDENTIFYING CONTACT-INDUCED CHANGE	11
1.2.4 CAVEATS	14
1.3 DESCRIPTION OF FIELD SITE	14
1.3.1 GENERAL INFORMATION	14
1.3.2 SOCIOLINGUISTIC SETTING IN THE VILLAGES	17
1.3.3 THE CONSULTANTS	19
1.4 OUTLINE OF THE THESIS	21
CHAPTER 2 THE DOLGAN PEOPLE	25
2.1 GENERAL INFORMATION	25
2.2 GEOGRAPHICAL ENVIRONMENT	28
2.2.1 ECOLOGY, FLORA AND FAUNA	28
2.2.2 SETTLEMENTS AND CITIES	30
2.3 HISTORY	31
2.3.1 PREHISTORY OF THE TAIMYR (UP TO 1638)	31
2.3.2 RUSSIAN EMPIRE (1638 - 1917)	32
2.3.2.1 ACCESS TO WEST SIBERIA	32
2.3.2.2 THE VILYUY AND LENA BASIN	33
2.3.2.3 THE TAIMYR PENINSULA	36
2.3.3 SOVIET PERIOD (1917 - 1989)	42
2.3.4 POST SOVIET PERIOD (1989 - PRESENT)	47
2.4 ETHNOGRAPHY	48
2.4.1 RUSSIAN EMPIRE (1638 - 1917)	48
2.4.2 SOVIET UNION (1917 - 1989)	51

2.5 LANGUAGE	59
2.5.1 LANGUAGES ON THE TAIMYR	60
2.5.2 DOLGAN: A DIALECT OR A LANGUAGE?	62
2.6 GENETIC COMPOSITION OF THE DOLGANS	65
2.6.1 BASIC CONCEPTS.....	65
2.6.2 MTDNA ANALYSIS	69
2.6.3 Y-CHROMOSOME ANALYSIS.....	71
2.6.4 INTERPRETATION.....	74
2.7 SUMMARY	77
CHAPTER 3 LANGUAGE CONTACT THEORY	81
3.1 THE FIELD OF LANGUAGE CONTACT STUDIES	81
3.1.1 DEFINITION AND BRIEF HISTORY OF THE FIELD.....	81
3.1.2 STRUCTURAL CONSTRAINTS	84
3.1.3 PSYCHOLINGUISTIC CONSTRAINTS	87
3.1.3.1 SECOND LANGUAGE ACQUISITION	88
3.1.3.2 BILINGUAL ACQUISITION.....	91
3.1.3.3 BILINGUALISM IN LANGUAGE CONTACT THEORY	92
3.1.4 SOCIAL FACTORS.....	95
3.1.4.1 INTENSITY OF CONTACT	96
3.1.4.2 EMBLEMATICITY (ROSS).....	100
3.1.4.3 SOCIO-SPATIAL FUNCTION AND ATTITUDE (THURSTON, ANDERSEN, WRAY & GRACE	102
3.1.5 FROM INNOVATION TO CHANGE.....	104
3.1.5.1 NETWORK TIES (MILROY AND MILROY)	105
3.1.5.2 AGE.....	107
3.1.6 SUMMARY	108
3.2 THE ROLE OF THEORY IN LANGUAGE CONTACT STUDIES	110
3.3 TERMINOLOGY	113
CHAPTER 4 LEXICON	117
4.1 INTRODUCTION.....	117
4.2 LEXICAL CHANGE AND LANGUAGE CONTACT	118
4.3 METHODOLOGY AND CATEGORIES OF CLASSIFICATION	121
4.3.1 METHODOLOGY	121
4.3.2 CATEGORIES OF CLASSIFICATION.....	123
4.4 RESULTS	125

4.4.1 QUANTITATIVE ANALYSIS	125
4.4.1.1 ONOMASIOLOGICAL PERSPECTIVE	126
4.4.1.2 TYPES OF DIFFERENCE.....	128
4.4.2 QUALITATIVE ANALYSIS.....	131
4.4.2.1 SEMANTIC CHANGE	131
4.4.2.2 REPLACEMENT	144
4.4.3 SUMMARY	158
4.5 INTERPRETATION OF RESULTS AND CONCLUSION.....	160
CHAPTER 5 REGULARISATION OF PARADIGMS	165
5.1 INTRODUCTION.....	165
5.2 REGULARISATION OF NOMINAL PARADIGMS.....	166
5.2.1 DESCRIPTION OF THE PHENOMENON	166
5.2.2 REANALYSIS	168
5.2.3 SAKHA NOMINAL DECLENSION	169
5.2.3.1 RELATIONAL NOUNS AND REFERENTIAL NOUNS	169
5.2.3.2 NOUNS IN PHRASES OF LOCATION.....	170
5.2.3.3 REGULAR STEMS AND UNSTABLE STEMS	172
5.2.4 DOLGAN EQUIVLENTS TO THE SAKHA FORMS.....	176
5.2.4.1 RELATIONAL NOUNS	176
5.2.4.2 REFERENTIAL NOUNS.....	188
5.2.4.3 EARLIER EXPLANATIONS IN THE LITERATURE	190
5.3 REGULARISATION OF THE PARADIGM OF AUXILIARY VERB <i>E</i> - 'BE'	195
5.3.1 DESCRIPTION	195
5.3.2 DIRECTION OF CHANGE AND EARLIER ACCOUNTS IN THE LITERATURE	198
5.4 DISCUSSION	199
5.4.1 PRINCIPLES UNDERLYING PARADIGM REGULARISATION AND REANALYSIS	200
5.4.2 REGULARISATION AND REANALYSIS IN DOLGAN EXPLAINED	202
5.4.3 THE POTENTIAL ROLE OF L1 AND L2 LEARNERS IN REANALYSIS	204
CHAPTER 6 HABITUAL.....	209
6.1 INTRODUCTION.....	209
6.2 HABITUAL IN DOLGAN AND SAKHA AND THEIR POSITION AMONG OTHER TURKIC LANGUAGES	210
6.3 THE HABITUAL IN <i>-A:ččI</i>	213
6.3.1 THE ORIGIN OF <i>-A:ččI</i>	213
6.3.2 USE OF <i>-A:ččI</i> IN DOLGAN AND SAKHA.....	215

6.3.3 DIFFERENCES BETWEEN DOLGAN AND SAKHA.....	217
6.3.3.1 -A:ččI vs. -A:ččIk	217
6.3.3.2 NEGATION	217
6.3.3.3 VERBAL, NOMINAL AND ADJECTIVAL USE	218
6.3.3.4 FREQUENCY OF USE	220
6.4 PROBING THE CAUSE OF THESE DIFFERENCES IN DOLGAN AND SAKHA	224
6.4.1 LANGUAGE-INTERNAL MOTIVATIONS.....	224
6.4.2 LANGUAGE-EXTERNAL MOTIVATIONS	225
6.4.2.1 MORPHOSYNTACTIC PROPERTIES OF HABITUAL IN TUNGUSIC LANGUAGES	225
6.4.2.2 FREQUENCY OF HABITUAL ASPECT IN TUNGUSIC LANGUAGES	230
6.5 INTERPRETATION AND CONCLUSION	234
CHAPTER 7 WORD ORDER	241
7.1 INTRODUCTION.....	241
7.2 WORD ORDER IN TURKIC LANGUAGES	242
7.3 WORD ORDER IN SAKHA	244
7.4 WORD ORDER IN DOLGAN	246
7.5 QUANTITATIVE COMPARISON OF WORD ORDER PATTERNS IN DOLGAN AND SAKHA	249
7.5.1 METHODOLOGY	249
7.5.2 RESULTS	251
7.6 DISCUSSION.....	254
7.6.1.1 LANGUAGE-INTERNAL MOTIVATIONS FOR WORD ORDER CHANGE	254
7.6.1.2 LANGUAGE-EXTERNAL MOTIVATIONSS APPLIED TO DOLGAN	256
7.6.2.1 LANGUAGE CONTACT AND WORD ORDER CHANGE	258
7.6.2.2 LANGUAGE ACQUISITION AND WORD ORDER CHANGE	260
7.6.3 WORD ORDER VARIATION IN DOLGAN EXPLAINED	262
CHAPTER 8 CLAUSE COMBINING	267
8.1 INTRODUCTION.....	267
8.1.1 DEFINING COORDINATION AND SUBORDINATION	267
8.1.2 TERMS AND DEFINITIONS	272
8.2 COORDINATION.....	273
8.2.1 COORDINATION IN SAKHA	273
8.2.2 COORDINATION IN DOLGAN.....	276
8.2.3 DIFFERENCES BETWEEN DOLGAN AND SAKHA.....	278
8.2.3.1 ABSENCE OF UONNA 'AND' IN DOLGAN	278
8.2.3.2 FREQUENT USE OF ONTON 'AND THEN' IN DOLGAN.....	282

8.2.3.3 THE USE OF GITTA 'WITH' IN INCLUSORY COORDINATION CONSTRUCTIONS	286
8.2.3.4 RUSSIAN COORDINATORS IN DOLGAN DISCOURSE	292
8.2.3.5 RANGE OF ACTIVELY USED COORDINATORS	293
8.2.4 DISCUSSION	295
8.3 SUBORDINATION	296
8.3.1 ADVERBIAL SUBORDINATION	297
8.3.1.1 PURPOSE	297
8.3.1.2 TEMPORAL RELATIONS	302
8.3.1.3 CONDITIONAL RELATIONS	305
8.3.1.4 REASON	306
8.3.2 RELATIVE RELATIONS	307
8.3.2.1 RELATIVE CLAUSES IN SAKHA	309
8.3.2.2 RELATIVE CLAUSES IN DOLGAN	311
8.4 THE USE OF RUSSIAN LINKING ELEMENTS	318
8.5 CONCLUSION	323
CHAPTER 9 DISCUSSION	327
9.1 DRAWING THE STRANDS TOGETHER	327
9.2 SUMMARY OF RESULTS	329
9.3 CONTACT WITH EVENKS	330
9.3.1 LINGUISTIC CHANGES	330
9.3.1.1 A NOTE ON PHONETICS	336
9.3.2 INTERPRETATION OF THE LINGUISTIC DATA WITHIN LANGUAGE CONTACT THEORY AND GENETICS	340
9.3.2.1 CONTACT BEFORE ARRIVAL ON THE TAIMYR	340
9.3.2.2 CONTACT ON THE TAIMYR	343
9.3.3 SUMMARY	348
9.4 CONTACT WITH RUSSIANS	349
9.4.1 LINGUISTIC CHANGES	350
9.4.2 INTERPRETATION OF CHANGES USING LANGUAGE CONTACT THEORY, SOCIO- HISTORICAL AND GENETIC INFORMATION	354
9.4.3 SUMMARY	358
9.5 THE USE OF DOLGAN/SAKHA AS A LINGUA FRANCA	358
9.6 CONCLUSION	361
CHAPTER 10 CONCLUDING REMARKS AND FUTURE OUTLOOK	363
REFERENCES	371

SUMMARY	389
SAMENVATTING.....	392
КРАТКОЕ СОДЕРЖАНИЕ.....	396
CURRICULUM VITAE.....	400

ACKNOWLEDGEMENTS

Reflecting on how many people contributed to the realisation of this project, it seems wrong to see just a single author on the front page of this book. While I obviously carry full responsibility for its content, this volume would simply not exist without the invaluable help of many others, who advised and supported me in uncountable ways.

First of all I want to thank Brigitte Pakendorf for her inspiring ideas, her commitment and for her absolute support. Not only did she introduce me to the Dolgan people and their intriguing history, but she also revealed to me new potentials of linguistic data when they are combined with genetics. The idea to combine these two research fields to gain insight into Dolgan prehistory is entirely hers, and the modest knowledge I acquired with respect to population genetics I owe to her patient explanations. I thank her for sharing her Sakha and Even data with me, for her instant and careful reviews of my work and for our fruitful discussions.

My gratefulness also goes to Maarten Mous, who accepted me as an external PhD student at the University of Leiden, and whose refreshing and insightful comments helped me keep the big picture in mind while finalising the details of this work.

Most essential for the accomplishment of this project has been the hospitality of the Dolgan and Sakha people and their willingness to share their language and daily life with me. My visits to the Dolgans and the Sakha have enriched my life in unparalleled ways, and without their kindness, collaboration and trust the entire enterprise would have been pointless. I warmly thank every single person I had the privilege to work with and who made my visits to the Taimyr possible. In particular, I would like to thank the following people for telling me their stories, for their collaboration in linguistic work and for their patience while initiating me in the Dolgan and Sakha languages: Tatyana Porotova, Vasilij Porotov, Svetlana Chuprina, Anna Timchenko, Nina Bettu, Vladislav Kokore, Maria Chardu, Irina Porotova, Afanasiy Fedosiev, Stepanida Bettu, Praskovya Kirgizova, Irina Antonova, Makar Antonov, Aksinya Chuprina, Diana Kudryakova, Nikolay Chuprin, Spiridon Kudryakov, Marina Antonova, Lyudmila Baykalova, Tamara Suvorova, Larisa Sotnikova, Ekaterina Bettu, Aleksandr Silkin, Evdokia Bettu, Anna

Barbolina, Elena Kirgizova, Lina Kirgizova, Ol'ga Makarova, Anastasia Rakhleeva, Aitalina Androsova, Boris Mandar, Valentina Khristoforova, Margarita Egorova, Nikolay Egorov. In addition I express my gratitude to Ol'ga Zhilkina and Ivan Zhilkin for the occasional apple they brought in from town and for their *banya*, which improved my personal well-being considerably.

I thank Avgusta Pavlovna, Nina Fedorovna, Natalia Fal'kova and Bella Vasilievna for accepting me into the villages, and most notably Oksana Kudryakova and Evdokim Kudryakov for hosting me in their *baloks* in the tundra. I am particularly indebted to Vera Rudinskaya, Inga Rudinskaya, Nina Koval'chuk, and Anna Bychkova, whose house was always open for me, and whose friendship and hospitality were more heartwarming than any Siberian stove could ever be.

The organisation of my fieldwork was greatly facilitated by our collaboration with Dom Narodnogo Tvorchestva in Dudinka and the Institute of Problems of Minority Peoples of the North in Yakutsk. I thank Lyubov' Popova, Tatyana Ruban, Nina Kudryakova and Larisa Bettu for dealing with the complex administrative procedures that are required to invite foreigners into the region, and Alexandr Maksimov at the border police in Dudinka for the necessary permissions to enter the villages in the Khatanga region. Furthermore, I am grateful to Innokentiy Machitov, Anatolij Savvinov and Marina Dmitrievna for helping with practical matters during my stay in Yakutsk and Dudinka.

Closer to home, I express my gratitude to the Max Planck Society for its generous funding and for creating optimal research conditions. On a more personal level, I thank my colleagues and friends at the Max Planck Institute for Evolutionary Anthropology in Leipzig who added irreplaceable intellectual and personal value to the time I spent at this institute. I am grateful to Peter Fröhlich and Hans-Jörg Bibiko for technical support and for their help with the creation of databases, Silvio Tüpke and Marike Schreiber for transforming my cartographic ideas into pretty maps, Roger Mundry, Michael Danneman and Matthias Urban for patiently explaining statistics to me and for choosing the right statistical models to represent my data, Jana Neuwirt for coding Evenki texts, and Martin Haspelmath, Malcolm Ross and Frank Seifart for discussing linguistic issues with me. The Max Planck research group for Comparative Population Linguistics proved an ideal context to carry out this research, and I thank all its members for their inspiring presence. In particular I thank Mark Whitten for sharing the results of his genetic analyses of the Dolgans with me, and for allowing me to use them in this thesis. My deep gratitude goes out to Bethwyn Evans, for her professional and

personal encouragement throughout the entire project. I thank her for her critical reading of my work, for correcting my English, for many insightful conversations, regardless whether they took place in the Irish hills, in the office, or at outings to the Leipzig moose. I thank my office mate Chiara Barbieri for spending an incredible number of hours with me in a small room without complaining, and Natasha and Katya Aralova for translating the summary of this work into Russian.

Finally I thank my close friends and my family. You bore with me despite geographical distance and my long periods of absence. Only a dedicated volume could express the deep gratitude I feel for your unconditional support, and would do justice to the recognition you deserve.

REFERENCE INFORMATION

ABBREVIATIONS AND GLOSSES

1	- first person	EXIST	- existential
2	- second person	F	- feminine
3	- third person	FUT	- future
ABL	- ablative	GEN	- genitive
ACC	- accusative	GNR	- generic
ADJZR	- adjectivizer	HAB	- habitual
ADVLZR	- adverbializer	IMP	- imperative
AG	- agent	IMPV	- imperfective
AOR	- aorist	INDEF	- indefinite
AUX	- auxiliary	INF	- infinitive
CAUS	- causative	INTNS	- intensifier
CIR	- circumstantial	INST	- instrumental
COLL	- collective	LAT	- lative
COM	- comitative	LOC	- locative
COND	- conditional	M	- masculine
CONTR	- contrastive	MOD	- modal
CV	- converb	MULT	- multiplicative
DAT	- dative	N	- neuter
DECL	- declarative	NLZR	- nominalizer
DEF	- definite	NONEXIST	- nonexistential
DEM	- demonstrative	NONFUT	- nonfuture
DET	- determiner	NEG	- negation
DER	- derivation	ORD	- ordinal
DIM	- diminutive	PART	- partitive
DISTR	- distributive	PASS	- passive
DUR	- durative	PF	- perfective
EMPH	- emphasis	PST	- past
EP	- epenthetic	PL	- plural
EXCL	- exclusive	POSS	- possessive

PRED	- predicative
PROG	- progressive
PRON	- pronoun
PROP	- proprietive
PRS	- present
PRT	- particle
PURP	- purposive
PTC	- participle
Q	- question
REL	- relative
RFL	- reflexive
R	- from Russian
RECP	- reciprocal
RES	- resultative
SG	- singular
SIM	- simultaneous
SQ	- sequential
VBLZR	- verbalizer

NOTE: case followed by person marking represents the possessive form of that particular case, e.g. ACC.3SG is the third person singular possessive form of the accusative.

DOLGAN TRANSCRIPTION SYMBOLS AND THEIR IPA VALUE

<i>Symbol</i>	<i>IPA value</i>	<i>Symbol</i>	<i>IPA value</i>
a	- a	n	- n
b	- b	ń	- ɲ
č	- tʃ	o	- o
d	- d	ö	- ø
e	- e	p	- p
g	- g	ʏ	- ɤ
h	- h	r	- r
i	- i	s	- s
ï	- u	t	- t
dʒ	- dʒ	u	- u
d'	- d'	ü	- y
k	- k	v	- v
l	- l	χ	- q
m	- m	:	- length

ETHNONYMS AND LINGUISTIC CLASSIFICATION OF SIBERIAN PEOPLES

<i>Ethnonym</i>	<i>Language</i>	<i>Language family</i>	
Dolgan	Dolgan	Northern Turkic	Altaic
Sakha (Yakut)	Sakha (Yakut)	Northern Turkic	Altaic
Khakas	Khakas	Northern Turkic	Altaic
Evenk	Evenki	Northern Tungusic	Altaic
Even	Even	Northern Tungusic	Altaic
Negidal	Negidal	Northern Tungusic	Altaic
Nanaj	Nanaj	Southern Tungusic	Altaic
Udighe	Udighe	Southern Tungusic	Altaic
Nganasan	Nganasan	Samoyedic	Uralic
Nenets	Nenets	Samoyedic	Uralic
Enets	Enets	Samoyedic	Uralic
Russian	Russian	Slavic	Indo-European
Chukchi	Chukchi	Chukot	Chukotka-Kamchatkan
Yukaghir	Yukaghir	Yukaghir	

MAPS

- Map 1: Current distribution of indigenous peoples of the Taimyr and neighbouring regions
- Map 2: Dolgan villages and towns on the Taimyr Peninsula
- Map 3: Distribution of indigenous people and Khatanga Trading Way on the Taimyr Peninsula and neighbouring regions in early 18th century
- Map 4: Peoples and locations where genetic samples were collected

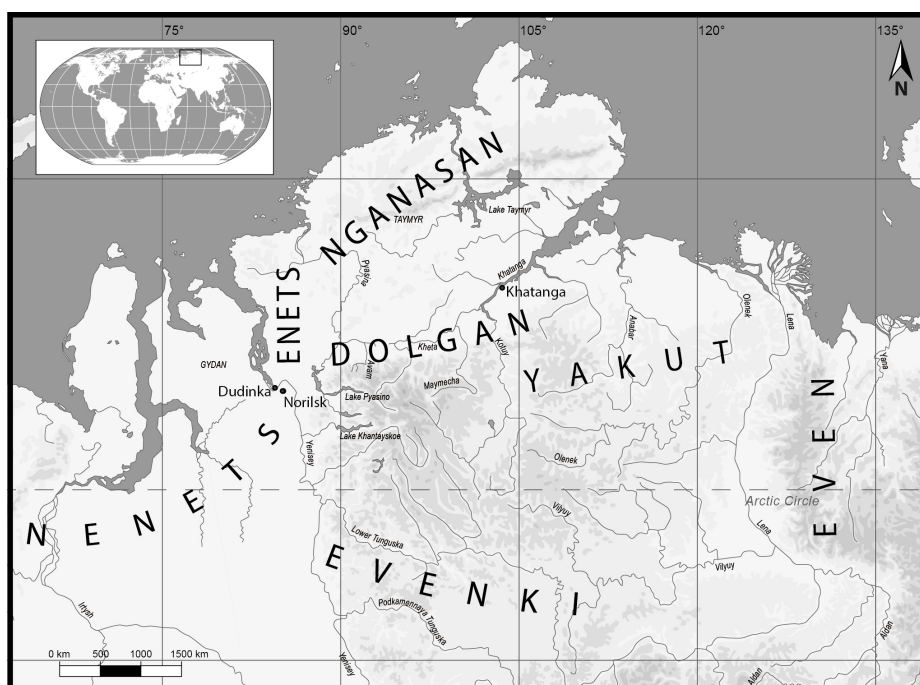
1.1. AIMS OF THE THESIS

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 2010¹).



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 ancestry². 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 indispensable.

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 ELAN³ (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 Rules⁴.

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 laptop⁵. 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 1,500⁶ 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 dialects¹¹. 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.

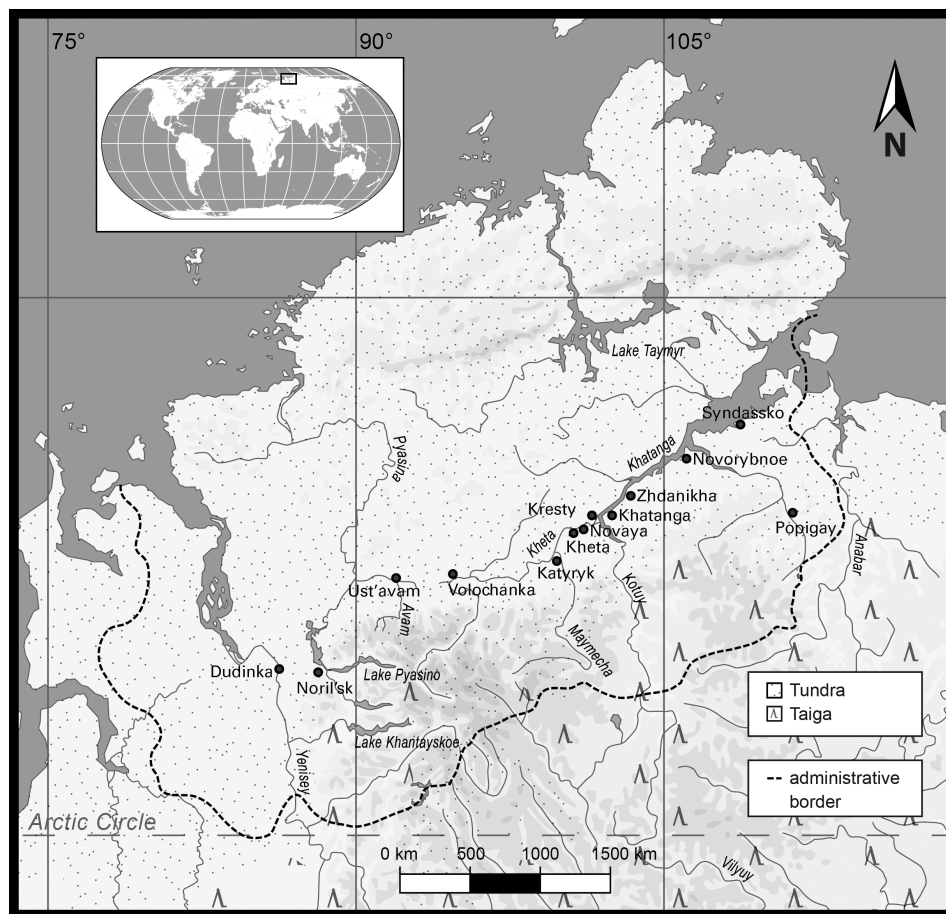
Second, there are differences in the nature of the data under comparison. While spoken corpora were available for Dolgan and Sakha, this was not the case for the languages they were compared with. To my knowledge, there is no accessible corpus of spoken annotated Evenki or Nganasan texts, which would 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 1.1¹².

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 (Voločanka, 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 1.2¹³.

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

Location of recording	Initials	Age	Gender
Voločanka	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/ Levinskie Peski	ESB	76	F
	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.

2.1 GENERAL INFORMATION

The Dolgan people are the northernmost Turkic-speaking population in the world. Their territory is situated entirely above the Arctic Circle, and comprises the Taimyr Peninsula and certain parts of the neighbouring Anabar district in the Republic of Sakha (Yakutia) (Savvinov 2005: 7, Ventsel 2005: 6, see Map 1 in Section 1.1). The etymology of their ethnonym has multiple potential interpretations. It is certain that ‘Dolgan’ has Tungusic origins, as it occurs repeatedly as the name of a Tungusic clan in different parts of Siberia, varying from Dolgan to Dulgan or Dulgaan. The most plausible interpretation is that it comes from the root *dul-* ‘middle’ in Evenki and Even, denoting inhabitants of the middle of the river, as opposed to those upriver and downriver.

The self-identification of the Dolgans on the Taimyr does not always match this official label. In all linguistic and ethnographic sources it is reported that they self-identify as *tīa kihite*, or *tīa*, the Dolgan equivalent of ‘tundra person’ or simply ‘tundra’. In addition, they may identify as *tege*, the Evenki word for ‘human’ or as ‘Sakha’¹ (Dolgikh 1963: 150). This inconsistency in (self)-denotation may have to do

¹ During my trips in 2008, 2009 and 2010, I did not encounter people who self-identified as either *tīa kihite* or *tege*. This may have to do with my being a foreigner, for whom people use their official label, or it might be that this characterisation is becoming old-fashioned. Most people self-identified as Dolgan

with the relatively recent formation of the Dolgans as a separate ethnolinguistic group, and with the concurrent changes in naming during this process. It may also be related to an existing view that Dolgan as an ethnolinguistic unit was created by outsiders (ethnographers, politicians) and does not reflect an internally coherent ethnic group (Anderson 2000: 74, 78, see also Section 2.4.2).

The Dolgans are the most numerous group of indigenous people on the Taimyr Peninsula: according to the most recent counts, 5,517 Dolgans were living in the Taimyr Autonomous District, which was renamed the Taimyr Municipal District in 2007. This corresponds to 54% of the entire indigenous population of the region, while the Nenets, Nganasan, Evenks and Enets together make up the remaining 46%². The Sakha, who constitute the sixth ethnic group of the wider region, are primarily associated with the neighbouring Anabar district to the east of the Taimyr.

The Dolgans have not always been in the numerically dominant position they occupy today. In fact, they are often described as the ‘youngest’ ethnic group on the Taimyr Peninsula, whereby ‘youngest’ refers to their formation as a separate ethnolinguistic group, and not to the first time the clan name appears in ethnographic accounts. It is commonly recognised that the people who identify as Dolgan today are a mix of Tungusic (Evenk), Turkic (Sakha) and Slavic peoples (Russian Tundra Peasants) (e.g. Dolgikh 1963: 93, Ubryatova 1985: 5, Forsyth 1992: 56, Slezkine 1994: 102, Anderson 2000: 9, 85). However, there is no agreement as to the moment these different ethnic groups began to consolidate into a new community, how exactly this happened, and which factors motivated this development. Estimates vary from the 17th century (Ubryatova 1985: 8, Stachowski 1996: 129) to as recently as the 20th century (Dolgikh 1963: 137). This wide time span can be explained by the fact that the term ‘Dolgan’ has been given different interpretations by ethnographers, historians and politicians over time, and by the concurrent administrative changes on the Taimyr Peninsula, which promoted or demoted the recognition of the Dolgans as an ethnolinguistic group.

and some older people as Sakha, which reflects the fact that this was their official name between 1935 and 1961 (see Section 2.4.2 for more details). However, they recognised *tia kihite* as a way to refer to themselves as an ethnolinguistic group.

² The numbers for all indigenous ethnic groups (confirmed in 2008) in the Taimyr Municipal District are as follows: Dolgan - 5517, Nenets - 3468, Nganasan - 749, Evenks - 270, Enets - 168, other - 27 (cited from http://www.taimyr24.ru/about/index.php?SECTION_ID=122&ELEMENT_ID=646 accessed on January 18th 2012). Altogether, the indigenous population comprises 27% of the entire population on the Taimyr.

The language of the Dolgan people is also called Dolgan. Together with Sakha, it belongs to the northeastern branch of the Turkic language family, and it can be divided into two dialects: the *üöhegi* or upriver dialect, and *allaragi* or downriver dialect. The upriver dialect is spoken in the villages Khantayskoe Ozero, Ust' Avam, Volochanka, and Katyryk, the downriver dialect in Novaya, Kresty, Khatanga, Zhdanikha, Novorybnoe, Syndassko and Sopochnoe (see Map 2 in Section 1.3.1). The people in Kheta, where I conducted part of my fieldwork, characterised their Dolgan variety as 'a mixture' of both dialects. The dialects are very similar and differ predominantly in terms of certain lexical items and certain phonetic features.

Linguistically, Dolgan is very closely related to Sakha, the main language spoken in the Republic of Sakha (Yakutia). The languages are largely mutually intelligible with misunderstandings caused predominantly by differences in lexicon, pronunciation and rate of speech. It is easier for Sakha speakers to understand Dolgan than the other way round. This is probably due to the wider geographical distribution of Sakha as well as to its higher prestige and greater prominence in media (radio, newspaper). The estimated number of Dolgan speakers is 1,054 or 13.4% of the Dolgan population (Russian census 2010). Bilingualism is omnipresent and the Russian language is gaining ground quickly. As an illustration, the Russian census of 2002 still reported that 67% of the Dolgan population speak Dolgan. Although the significance of these numbers should be evaluated with some scepticism, it is certain that the number of speakers is declining rapidly.

While everybody has native or near-native command of Russian, four very broad categories can be observed within the Dolgan population with respect to their linguistic dominance. People over 65 are bilingual, but mostly dominant in Dolgan. Those between 45 and 65 show true bilingualism and have equally good command of both languages. For people younger than 45, Russian is clearly the dominant language, and under 25 it is hard to find fluent Dolgan speakers at all. It goes without saying that there is a large amount of individual variation within these categories. For example, within the 45-65 category, Dolgan proficiency generally declines as people get younger. However, these categories illustrate the general process of an on-going language shift to Russian. This picture represents the language situation in all villages except Syndassko and Sopochnoe, where everyone over five years old is bilingual, but where Dolgan still is the dominant language for everyday use, and children still learn it as their mother tongue.

In order to understand the diverse opinions about the Dolgan people and their language, it is necessary to be aware of the historical, ethnographic and genetic accounts that each give a different perspective on their fascinating history. As will become clear in the remainder of this chapter, it seems that such accounts have not only described but also actively shaped the Dolgans as an ethnic group. Without pretending to be exhaustive, this chapter is intended to provide the essential background information from these three perspectives. After a brief description of their natural environment, the emergence of the Dolgan people will be embedded in historical (Section 2.3), ethnographic (Section 2.4), linguistic (Section 2.5) and genetic (Section 2.6) contexts.

2.2 GEOGRAPHICAL ENVIRONMENT

2.2.1 ECOLOGY, FLORA AND FAUNA

The primary area of residence of the Dolgan people today is the Taimyr (Dolgan-Nenets) Municipal District³. The label of this administrative unit reflects its geographical location (Taimyr), as well as the names of the two largest indigenous ethnic groups that currently inhabit this territory (the Dolgans and the Nenets). It is divided into four administrative districts (Dudinka, Khatanga, Dikson and Karaul) and the administrative centre is in the city of Dudinka. The entire district, which consists of the Taimyr Peninsula and adjacent areas to the south and east, is located north of the Arctic Circle and includes the northernmost tip of the Eurasian mainland, Cape Chelyuskin. It covers 879,900 square kilometres, which roughly corresponds to two and a half times the size of Germany (357,021 square kilometres), with a population density of 0.045 persons per square kilometre (cf. 229 for Germany)⁴. This vast area is characterised by two main ecological zones: forest tundra in the south, and tundra, or Arctic desert, in the north. The boundary between these ecological zones, which coincides with the tree line, runs right across the peninsula and plays an important role in the distribution and movement of humans and animals in the region.

The forest tundra, which is a transitional zone between the dense forest of the taiga further south and the moon-like landscape of the treeless tundra in the

³ Таймырский (долгано-ненецкий) муниципальный район.

⁴ Numbers are taken from the website of the Taimyr: www.taimyr24.ru, accessed on January 18th, 2012.

north, is characterised by the presence of larch trees, willow weed and dwarf birches, interspersed by patches of barren tundra. North of the tree line, the tundra stretches all the way north to the Arctic Ocean. Here the landscape looks entirely devoid of trees and the severe climate only allows for the growth of mosses, lichens and some polar willows (Ziker 1998: 63, Anderson 2000: 11). These willows are hard to recognise as trees, as they have adapted to the severe climate by growing almost horizontally, and they serve as a means of orientation for people travelling in the endless snowy tundra in winter (personal observation).

Water is another prominent feature of the area, in winter in the shape of snow and ice, in summer in the shape of majestic rivers, in particular the Yenisey and Kheta (turning into Khatanga), which cross-cut the peninsula from south to north and from west to east respectively. In addition, summer reveals a myriad of lakes, puddles and swamps, which rapidly emerge as the rays of the sun gain in strength and cause the solid, frozen landscape to melt. The combination of the melting snow on the surface and the permafrost below prevents the water from being absorbed into the soil and thus provides ideal conditions for vast quantities of migratory (water)birds (geese, ducks, loons, storks, falcons (Ziker 1998: 67), and an even more overwhelming quantity of thirsty mosquitoes.

In addition, this region is home to rock ptarmigans, lemmings, wolves and bears (brown bears in the south, polar bears in the north), but most important to the indigenous people are the herds of wild reindeer, which are reported to be among the largest in the world (Ziker 1998: 67). Within living memory, reindeer have always played a crucial role in the maintenance of human life in the area, as they provide a reliable source of food, clothing, transport and even building material in a natural environment that otherwise provides rather unfavourable living conditions. The reindeer were also an important reason for the widespread nomadism in the area. In fact, none of the indigenous peoples was originally sedentary. The unfeasibility of agriculture in this climate has led to a longstanding symbiosis of man and reindeer, where humans followed the migrating reindeer according to the rhythm of the seasons: north in summer and south in winter. However, this situation has been changing over the last 80 years or so, with increasing industry in the 1950's and the forced settlement in villages in the 1970's having dramatic consequences for animals and people (see Sections 2.3.3 and 2.3.4).

2.2.2 SETTLEMENTS AND CITIES

While the Dolgans remain mostly dependent on reindeer for food and to some extent for fur and transport, after the forced settlement of the 1970's most of them live permanently in villages. Even the families who spend most of their time in the tundra with the reindeer herd are officially registered in a village and have a house. Today the Dolgans live in ten villages in the Taimyr Municipal District, as well as in the towns of Dudinka and Khatanga. They are positioned in a line across the Peninsula from west to northeast⁵, linking the two larger towns of Dudinka and Khatanga and stretching beyond them. This distribution is by no means a coincidence. The line of villages roughly coincides with the tree line as well as with the Kheta and Khatanga rivers, which have provided a corridor for the transportation of people and goods for centuries, and which became known as the Khatanga Trading Way (see Sections 2.3.2.3 and 2.4.2 for details).

In the villages, which vary in size from roughly 400 to about 600 inhabitants, the Dolgan people constitute the absolute majority of the population. Only in Khantayskoye Ozero the population is mixed with Evenks, and in Ust' Avam, Volochanka and Novaya, Dolgans share the village with Nganasan people. It is worth noting that in Novaya the Nganasan people have adopted the Dolgan language, whereas in the two other villages the languages have been kept separate. The number of Russians can normally be counted on one hand, and typically they occupy positions in administration or are merchants. In the bigger towns of Dudinka and Khatanga the ethnic composition of the population is more heterogeneous, including Russians as well as people with other ethnic backgrounds (e.g. Khakasians, Ukrainians, Azerbaijanis). Nonetheless, the proportion of Dolgans is significant in the towns as well, particularly in Khatanga. Despite the problematic infrastructure there is quite a lively movement between villages and towns. Many young people come to Khatanga and Dudinka to study and many of them stay there after finishing their education. This in turn attracts more people from the villages who come over to visit their relatives in town or to do shopping and get supplies.

⁵ From west to east, these villages are Khantayskoye Ozero, Ust' Avam, Volochanka, Katyryk, Kheta, Novaya, Kresty, Zhdanikha, Novorybnoye, Syndassko and Sopochnoye (previously Popigay).

2.3 HISTORY

2.3.1 PREHISTORY OF THE TAIMYR (UP TO 1638)

Our recorded knowledge of the history of the indigenous people of Arctic Siberia begins only in the 17th century CE, when the Russians expanded their empire to include the vast unexplored lands behind the Ural Mountains. In order to map the area, they began to note down information on the inhabitants and their distribution across the territory. The main goal was to facilitate the collection of *yasak*, or tribute, which was extracted from the indigenous population in the form of furs. Obviously the prehistory of human occupation of the Taimyr extends much further back in time. However, since we have no means to physically travel back in time, the only available source of information about this period are archaeological findings, which at best can provide a patchy picture of the distant past.

The earliest evidence of human presence on the Taimyr Peninsula goes back to at least 7,000 years before CE. This estimate was made on the basis of bronze objects and crucibles for their production, which were discovered near the Volochanka River in 1967, and were later associated with east Siberian Mesolithic sites (Khlobystin 1972, Troitskiy 1987: 20, Khlobystin and Gracheva 1993, cited in Ziker 1998: 69, Denisov 2008: 8). It is known that in this period of time the climate at these latitudes was warmer than it is today, but no information has been preserved about the people who produced the objects. In contrast to these earliest discoveries, later ceramic objects dated 4,000 to 2,500 before CE show influences from east as well as west Siberian traditions, suggesting contact between people with different cultural traditions (Khlobystin and Gracheva 1993, cited in Ziker 1998: 69, Denisov 2008: 9). The assumption is that the earliest inhabitants of the Taimyr Peninsula were hunter-gatherers, related to the ancestors of today's Yukaghir, Chukchi and Inuit (Troitskiy 1987: 20 in Ziker 1998: 70).

In the 2nd to 4th century CE Samoyedic populations migrated north and eastwards and entered Arctic Siberia, including the Taimyr. The main incentive appears to have been the progression of the warfaring Hun tribes who conquered the Siberian south and who compelled the original population to escape to the north. It is presumed that these Samoyedic and later also Tungusic people eventually merged with the people already present on the Taimyr Peninsula. The migration of Samoyedic people into the Taimyr region continued in the period between the 10th and 15th centuries. They moved from west to east, where they

encountered the Tungusic groups, and spread both their language and culture across the Peninsula (Ziker 1998: 71).

In the 17th century the first Russians set foot on the current territory of the Dolgan people and this is when recorded history begins. However, at that time the presumed ancestors of the Dolgan were still living in the area between the Lena and Vilyuy Rivers in what is the present day Republic of Sakha (Yakutia). Therefore, in order to understand the history of the Dolgan people, two regions are of main importance: the area of the Lena and Vilyuy rivers where the ancestors of the Dolgan came from, and the Taimyr Peninsula, where they live today. Since the recorded history of the Dolgan is so closely intertwined with the presence of the Russians, first a brief sketch will be given of the penetration of the Russians into Siberia to illustrate the general climate in which the first encounters between Russians and indigenous people took place. After that the focus will move to the two geographical areas mentioned above.

2.3.2 RUSSIAN EMPIRE (1638 - 1917)

2.3.2.1 ACCESS TO WEST SIBERIA

The main motivation for the eastward expansion of the Russian empire into Siberia was the acquisition of fur. Besides a general curiosity about the unknown, which is deep-rooted in the human mind and drives most explorations, this “treasure of the land of darkness” (Slezkine 1994: 12) was the driving force for many Russians to risk their lives and conquer the vast territory behind the Ural Mountains. However, while its population may have been sparse, Siberia was anything but an empty land. At the time of the Russian conquest, Siberia was home to many different indigenous groups, speaking different languages, who had not exactly been waiting for the Russians to enter their hunting grounds and consequently did not receive them with joy. As much as these mysterious inhabitants were a danger to the Russians, they were also indispensable. After all, the indigenous people knew much better where to find and how to trap the sought-after sables, squirrels and foxes, and thus how to secure the fur for the future coats of people in Europe and Central Asia.

In theory, the assistance of these people was to be obtained voluntarily or at least in a non-violent manner, but reports of the actual procedures show that this resolution was easily abandoned if the approach did not have the desired effect for

the Russians. If the colonisers succeeded in subduing the indigenous Siberians, these would be registered as *yasak* (or tribute) people, which meant that they were obliged to supply a certain amount of fur to the tsar each year 'for ever and ever', in return for the tsars' 'protection' (Slezkine 1994: 15).

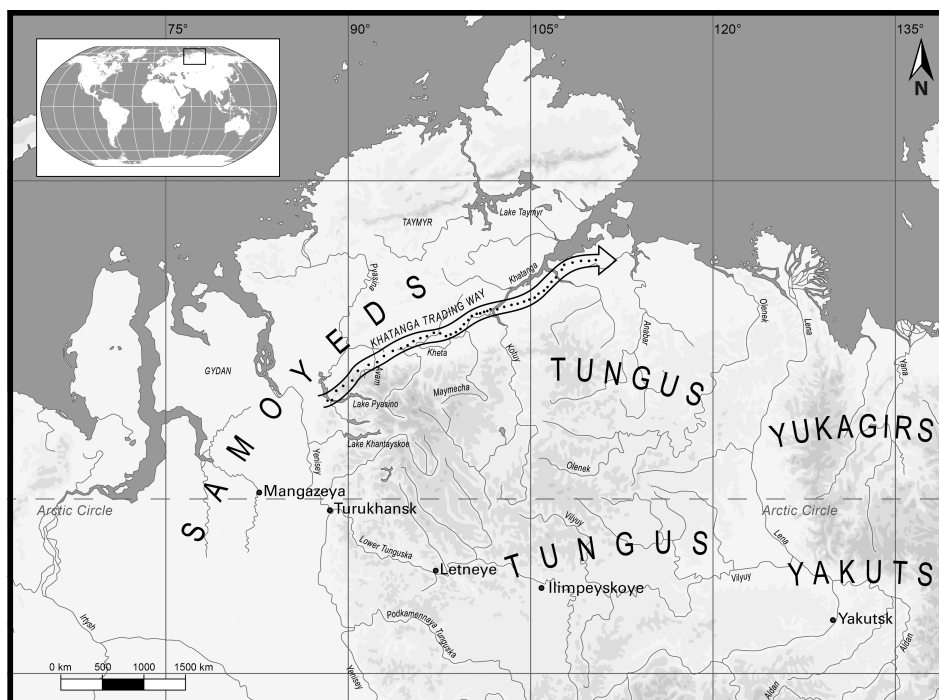
The fur trade is of principal importance because it had a dramatic impact on the lives of all indigenous Siberian peoples, as well as on the lives of the Russians themselves. The intensification of hunting practices, newly introduced diseases, as well as attempts to escape the tsar's 'exalted hand', which now extended deep into the Siberian taiga to tap into its rich resources (Gurvich 1966: 49 cited in Slezkine 1994: 21), led to new migrations, to new contacts, and for many populations to dramatic changes in their numbers and their manner of subsistence.

A key date in this development is the year 1552, when the khan of the Kazan Tatars was driven away, and the capital Kazan was conquered. From then on, the area between the Volga and the Ural Mountains belonged to Russia. This opened up the markets of Central Asia and the Middle East for the outflow of furs, which consequently led to an increase in demand (Forsyth 1992: 40, Slezkine 1994: 12). The real breakthrough in the Russian conquest of Siberia is typically considered the period of 1581-1582 when a Cossack army, headed by the Novgorodian Cossack Yermak Timofeevich, crossed the Ural Mountains and defeated Kuchum, the khan of Sibir, who had so far ruled the area (Slezkine 1994: 12-13, Ziker 1998: 71, Forsyth 1992: 30). This freed the way for hundreds, and later thousands, of trappers, mercenaries, soldiers and Cossacks to explore the immense stretch of land east of the Ural Mountains and all the way to the Pacific Ocean.

2.3.2.2 THE VILYUY AND LENA BASIN

In terms of the Russians securing access to the two areas that are significant in the history of the Dolgans, one event is of crucial importance. This is the founding of the fortress of Mangazeya in 1601 (see Map 3), which served as a springboard for the expansion of Russian power further east towards the Yenisey and Lena Rivers, as well as for the exploration of the Arctic regions of Siberia including the Taimyr Peninsula and the rivers Kotuy, Anabar and Olenek (Forsyth 1992: 57). Named after the local Samoyed tribe Mongkansi (Forsyth 1992: 36) this fortress thrived and quickly developed into a town that eventually became the capital of Central Siberia.

On their eastward journey from Mangazeya, the Russians soon reached the Yenisey River where they founded the fortresses of Turukhansk (1604) and Khantaysk (1620) (Forsyth 1992: 36). Travelling further up the Lower Tunguska and over land, they reached the great Lena River in the 1620's. Before the arrival of the Russians, this area was dominated by Tungusic people, whose territory extended eastward all the way to the Pacific Ocean and southeast into Mongolia and Manchuria. They shared this vast area with only two other ethnic groups, the Buryats at Lake Baikal, and the Sakha (or Yakut) people who at the time populated the confluence of the Lena and Vilyuy Rivers (Forsyth 1992: 48). Thus in this area Tungusic-speaking Evenk clans resided in the vicinity of the Turkic-speaking Sakha, and, as will be shown below, this coexistence and the consequent encounters are of great significance for the formation of the ethnolinguistic identity of the Dolgan people.



Map 3: Distribution of indigenous people and Khatanga Trading Way on the Taimyr Peninsula and neighbouring regions in the early 18th century

Turbulent times followed the arrival of the Russians. On the one hand, all the indigenous people had to protect themselves from the Russians, their indomitable

hunger for fur, and their diseases. On the other hand, internal rivalries between Sakha clans, as well as between the Sakha and Tungus clans made it impossible for them to join forces effectively against the invaders. Despite multiple efforts to defeat the Russians, which continued over fifty years (in particular by the Sakha (Forsyth 1992: 60) resistance did not last. By 1630 the Russians had subdued the Sakha on the Lena River to pay *yasak* to Mangazeya, and in the years to follow they extended their web of forts to the Olenek river where they established themselves among the Tungusic Edyan clan and to other Tungusic and Sakha clans along the Vilyuy and Aldan Rivers (1634-1638) (Forsyth 1992: 60).

An additional source of discontent in the indigenous communities were the internal rivalries among Russians themselves, and the consequences these had for their *yasak*-extracting practices. As the occupation of Siberia advanced, boundaries of districts changed, which often led to conflicts about who was entitled to claim *yasak* from the indigenous population in each area. Not unexpectedly, such problems were often 'solved' by both parties stubbornly insisting on the same right, with the implication that the Tungusic and Sakha clans had to pay their tribute twice. It can be safely assumed that such doubled *yasak* extraction corresponded to at least doubled discontent among the people who had to deliver it.

These conflicts, the oppression by the Russians, a *yasak*-load too much to cope with, in combination with the consequent dramatic drop in fur-bearing animals in the area was the incentive for a number of Tungusic and Sakha clans to leave their homeland and move to more northern and presumably safer lands. They moved to the basins of the Anabar, Olenek and Kotuy rivers, where they encountered other Tungusic clans and Samoyedic people. Here the Sakha clans, who had traditionally led a pastoralist lifestyle, had to adapt to their new Tungusic environment, and exchanged pastoralism for fishing and wild reindeer hunting as their main mode of subsistence (Forsyth 1992: 63).

Thus, from the mid 17th century onwards the area called northern Yakutia and southern Taimyr today, came to be populated with a variety of ethnic groups who were driven away from their original hunting and herding grounds. While the migrations of the indigenous groups can be partly attributed to the arrival of the Russians, this was probably not the only driving force. Being nomadic or semi-nomadic, moving to new territories was nothing unusual for many native Siberians. In addition, certain groups were simply in the process of expanding, most notably the Sakha. Starting from the relatively small area between the Lena,

Aldan and Amga rivers around the 13th century, to which they had migrated from further south, they now occupy the immense territory of Yakutia and are the dominant ethnic group in northeastern Siberia (Wurm 1996: 969, 971, Pakendorf 2006: 335). As a reaction to the influx of people from the south, the Samoyedic Tavgi (or Nganasan) population that had so far occupied the southern Taimyr retreated further north.

In these years the name 'Dolgan' appears in the historical records for the first time. In a document dated August 6th 1638 a certain Petr Golovin and Efim Filatov are instructed to "found a stockade town and to turn the Siberian aliens into Russian citizens"⁶, including the Dolgans. In this document they are mentioned in a list of Tungusic clans, and are described as people "whom nobody governs" (Russkaya istoricheskaya biblioteka, 1875: 968, cited in Ubryatova 1985: 8). Archival documents reveal that by 1638 these Dolgan people were living on the Lena River between the lower Vilyuy and the Aldan (Ubryatova 1985: 8, Dolgikh 1963: 107) and that they numbered between 90 and 120 people (Dolgikh 1963: 107). By 1644 they had already moved to the upper Vilyuy to escape the double *yasak* they were forced to pay to Mangazeya as well as to the town of Yeniseysk, but unfortunately this turned out to be no solution to the problem. According to Ubryatova, the struggle with the Russian Cossacks continued for a few more decades, and as a result the Dolgan clan dispersed over a large area: some wandered off to the east where they mixed with Tungusic Even groups, others isolated themselves in the upper reaches of the Vilyuy River and gradually moved northwest to the Taimyr Peninsula. The exact years of these migrations will probably remain a mystery for most of the populations. However, for the Dolgans who moved to the Olenek River, the time of their migration can be reconstructed to the period between 1655 and 1678 on the basis of *yasak* records (Dolgikh 1963: 108).

2.3.2.3 THE TAIMYR PENINSULA

As mentioned above, the Taimyr Peninsula has been inhabited by humans for at least 9,000 years, but it is uncertain when the first Russians set foot on its territory. This is not without reason. From the moment the disclosure of new fur-

⁶ "... для строения острога и приведения в русское подданство сибирских инородцев" (Russkaya istoricheskaya biblioteka 1875: 968, cited in Ubryatova 1985: 8).

trapping lands was made public, not only the indigenous people but also the Russian explorers themselves were obliged to pay tax on this land to the tsar, which explains why such discoveries were often kept secret for some time, and so do not appear with correct dates in historical documents (Ziker 1998: 72). One point of orientation is the year 1601, the year Mangazeya was built. Since this fortress became the main base for the exploration of the Siberian Arctic, the first appearance of Russians on the Taimyr cannot have been too far removed from this date.

One thing we can be certain about is that the first Russians arrived in an ethnographic landscape that was rather different from the situation on the Taimyr today. As was mentioned in Section 2.3.2.2, in the first half of the 17th century the ancestors of the Dolgan were still living to the southeast of the Taimyr, and when the first explorers travelled up the Kotuy River, they encountered some Tungusic groups, but first and foremost Samoyedic (Nganasan) populations, who occupied the large territory of the west Siberian Arctic ((Troitskiy 1987: 30 in Ziker 1998: 75, Forsyth 1992: 36, Dolgikh 1963:107). In a similar fashion to the Tungusic and Sakha groups further south, these people were not pleased with the prominent presence of Russians and they repeatedly attacked Russian fortresses from at least 1604 till 1672. By that time, the situation of the Russians in Mangazeya had become so unbearable that they abandoned the town and transferred their administrative personnel to Turukhansk on the Yenisey (Forsyth 1992: 46).

The main reason for the Russian Cossacks, tax collectors and hunters to persist in the inhospitable environment of the Taimyr was, as in other areas of Siberia, to procure fur. However, for this plan to work, an entire network of supporting personnel had to be mobilised to provide the necessary conditions for survival. Consequently there were also priests, craftsmen, merchants, and peasants among the newcomers to the Taimyr, who were concentrated primarily along the Kheta and Khatanga rivers (Forsyth 1992: 42). They set up small stations along these major rivers at a distance of 10 kilometres apart all the way across the peninsula from Dudinskoe on the Yenisey in the west, to Khatanga in the east (Stern 2005: 292). This comparatively lively corridor of transport, exchange and trade attracted people from different ethnic origins and became established as the Khatanga Trading Way (or Khatanga trakt) during the late 17th and early 18th centuries (see among others Anderson 2000: 86-86, Stern 2005: 292, Stern 2009: 388).

In more southern regions of Siberia, the ‘peasants’ mentioned above were literal peasants. They constituted an ever-growing group of Russians who had come to Siberia primarily to develop agriculture. As their numbers increased, their use of the land encroached more and more on the hunting grounds of the indigenous population, and these conflicting interests led to confrontations between the two groups (Forsyth 1992: 64). In Arctic Siberia such dangers were rather limited since the climate does not allow for much agriculture to be practiced, but the equally disturbing Russian fur-trappers were called ‘government peasants’ nonetheless, which is how the term ‘Tundra Peasants’⁷ or ‘old settlers’ has become common usage (Troitskiy 1987: 54 in Ziker 1998: 78 Slezkine 1994: 97).

Until the 19th century, the number of Russian inhabitants of the Taimyr was very low. Fur hunters, tax collectors and merchants arrived regularly, but only a small number stayed and settled there permanently. This changed when in 1811 the governor of Tomsk initiated a settlement program in order to develop and improve the transport and communication system across the Taimyr Peninsula. Russian peasants were sent to the Arctic to cultivate the tundra, which obviously turned into a complete disaster. Having seen many of their fellows die, the only way for the remaining peasants to survive was to adopt the lifestyle of the native population. They adopted the cultural practices, beliefs and languages of the surrounding Sakha and Evenk tribes and intermarriage was common. Although some of them still identified as Russian, by the late 19th century most of them had become indistinguishable from the native population.

According to the Russian ethnographer Dolgikh, intermarriage was frequent not only between the Russians and the indigenous people, but also between members of different indigenous groups who inhabited the area around the Khatanga Trading Way (Dolgikh 1963: 125). This becomes apparent from Table 2.1, which presents an overview of all the registered marriages that took place on the Taimyr between 1727 and 1883. The table is taken from Dolgikh’s famous work ‘The origin of the Dolgans’, and specifies the ethnic origins of the members of each couple. Dolgikh adduces the large number of interethnic marriages in the region, among others, as an important development for the formation of the Dolgan people as a separate ethnolinguistic group since it literally blurred ethnic boundaries. The names of the ethnic groups in the table are taken directly from Dolgikh’s work. The Dolgan, Dongot, Edyan, Karanto, and Evenks are considered Tungusic clans;

⁷ Translated from Russian: затундренские крестьяне.

the Tundra Yakuts (referred to as T. Yakut in the table) and the Yessey Yakuts (indicated as Y. Yakut) are what I refer to as Sakha; the Tundra Peasants are of Russian origin and Samoyedic groups include the Nenets and Enets. Surprisingly, the Nganasan are not mentioned at all.

Table 2.1 Registered marriages on the Taimyr Peninsula between 1727 and 1883 (from Dolgikh 1963: 125)

Men	Women									
	Dolg.	Dong.	Edyan	Kar.	T.Yak.	T.Peas	Y.Yak	Evk.	Sam.	Total
Dolgan	25	1	-	1	27	18	4	8	1	85
Dongot	7	-	12	2	-	2	-	6	1	30
Edyan	-	10	-	-	-	-	-	2	-	12
Karanto	7	1	-	-	3	2	1	4	1	19
T.Yakut	38	-	-	-	100	38	3	1	3	183
T.Peas,	13	3	3	2	26	28	-	9	3	87
Y. Yakut	1	1	-	-	-	-	13	1	-	16
Evenki	3	5	3	4	3	2	2	15	1	36
Sam.	3	1	-	-	1	8	1	-	-	14
Total	97	22	18	9	160	96	24	46	10	482

This overview shows that out of 482 marriages only 181 (37.5%) were endogamous, i.e. between people of the same clan or ethnic group; the remaining 62.5% took place between people of different ethnic origins. With respect to the Russian Tundra Peasants, only 15.3% married another Tundra Peasant. Curiously, the Tundra Peasant men and women both married outside their own group very frequently and not, as might be expected, just the men due to the shortage of women, which existed due to the fact the majority of the Russian explorers had been men. In fact the women even topped the men with 70.8 % of their marriages being to a non-Tundra Peasant, whereas for the men this was only 67.8%. The ethnic groups they married into most frequently were for the women the Tundra Yakuts (39.6%) and the Evenk clan called Dolgan (18.6%). The same pattern applies to the men, who married Tundra Yakut women in 29.9% of the cases, and Dolgan women in 14.9%.

For the Tungusic groups (comprising the Dolgan, Dongot, Edyan, Karanto and Evenks), marriages with non-Tungusic people were not as overwhelming as it was for the Tundra Peasants to marry outside of their ethnic group, but they were still

very common. Of all 259 Tungusic marriages registered between 1727 and 1883, 44.8% took place between two Tungusic partners, whereas 55.2% married a non-Tungusic partner, who could be of Turkic (31.7%), Russian (20.8%) or Samoyedic origin (2.7%). As for the Tundra Peasants, there is no notable difference between the choice of partners for men or for women.

For the Turkic groups (including the Tundra Yakuts and the Yessey Yakuts), the pattern is very similar to the Tungusic groups: 43.44% of the marriages took place between two Turkic individuals, whereas in 56.6% a partner with a different ethnic background was found. Of these 56.6%, 30.7% married a Tungusic partner, 24% a Tundra Peasant, and 1.9% a Samoyedic individual. Considering these data, Dolgikh has good reason to believe that interethnic marriages were very common among the ethnic groups who inhabited the Taimyr Peninsula, and that as a result ethnic boundaries became less prominent among this particular assemblage of people.

Any impediments to interethnic marriages were smoothed out even more after the introduction of Christianity, which created an additional bond between the people who were baptised as opposed to those who were not. In the eyes of the Russians it was the only way for the indigenous people to lose their 'alienness' and to become part of Russian society (Slezkine 1994: 42-43). The first church was built in Khatanga in the first decade of the 18th century, and the Dolgans are reputed for being not only the first, but also relatively willing, to adopt this new religion. This is one of the reasons why they were later characterised by Russian ethnographers (e.g. Popov 1930 in Anderson 2000: 81) as 'avant garde' people of the Taimyr. The relative ease with which the Dolgans were converted is very different to the reluctance shown by other indigenous groups, as for example the Nganasan who actively practiced shamanism until the 20th century (Forsyth 1992: 178) and never really embraced Christianity at all. Nonetheless, official baptism did not mean that traditional religious practices were instantly abandoned. Christianity was expressed primarily by the adoption of Russian names and surnames, but traditional worshipping and shamanism remained vivid until the 20th century. The Soviet regime radically put an end to this after it 'unmasked' the shamans as exploiters and they were forced to stop their activities through repression or execution (Slezkine 1994: 226, Ziker 1998: 98).

Thus, encounters between Russians and native Siberians have taken place from the very beginning of the colonisation of Siberia. However, while the lives of most indigenous people began to change from the moment the Russians appeared,

the most fundamental transformations took place in the 20th century. In part, this had to do with the increased scale on which the Russian influence was exerted, but more importantly with a conceptual change in Russian ideology about the role of the indigenous people in Russian society, in particular in the Soviet Union. This ideological change percolated into the realms of politics and ethnography, which in turn led to radical transformations of society, dramatically affecting the lives of the indigenous people as well as of the Russians themselves.

If in the 17th and 18th century the prevailing opinion was that the native Siberians were inferior savages one needed to protect oneself from, the spirit of German romanticism of the 19th century changed their status into superior innocents that needed to be protected (Slezkine 1994: 73-74). Whether superior or inferior to the Russians, they had always been conceived of as principally different. This 'alien' status could have negative as well as some positive consequences for the Siberians. Thus, while there is clearly nothing advantageous in the fact that it was not considered immoral to exploit the indigenous population for the delivery of fur and services, in hindsight there were certain advantages in the fact that they were never forced to merge completely with Russian society. As long as they delivered their *yasak* on time, they could still more or less do what they wanted. Thus, this 'otherness' had always allowed for a certain distance and autonomy in that it justified the maintenance of their own way of life.

This situation changed radically with the establishment of the Soviet regime in the 20th century. The new ideology, which promoted progress and equality among all people, required unconditional participation in the building and realisation of a socialist state, regardless of ethnic background (Slezkine 1994: Ch. 6). To make this ambition a success, people had to be enlightened, educated and integrated, and this required a conceptual change from regarding the Siberian natives as 'aliens' to treating them as 'comrades'. Despite the good intentions behind this ideology, the imposed integration and the paternalistic decision to 'educate' the indigenous people and 'develop' them into full members of the Soviet society in a sense interfered more fundamentally with their traditional way of life than the initial Russian invasion had done, and for many people even destroyed it.

2.3.3 SOVIET PERIOD (1917 - 1989)

Key events started after the Revolution in 1917 and the following Civil War. In the initial period after the takeover of the Soviet regime, the new ideology of the Communist Party and following political measures must have been received as an improvement by the Siberian peoples. However, by the end of the 1920's ideas were taken to an extreme, leading to extreme consequences for the population, since the importance of the Soviet State as a whole began to overrule the importance of the individuals who had to live in it.

But the start looked promising. After three centuries of colonisation and institutionalised inequality⁸ the Bolshevik Party published in 1917 the 'Declaration of the Rights of the Peoples of Russia'. This document granted equality and sovereignty to all nationalities and thus changed the position of the Siberian natives fundamentally, at least in theory. According to Forsyth (1992: 241):

In Lenin's theoretical view differences of nationality were trivial compared with class divisions and allegiances, so that autonomy was simply a transitional stage towards centralisation (...)

whereby the final goal was the "eventual merging of all nations" (Lenin cited in Forsyth 1992: 241). This foregrounding of equality and sovereignty clashed most strikingly with the utterly unequal *yasak* relation, which had characterised the interaction between Russians and indigenous people so far. Therefore payment of tribute was abolished immediately in 1917. Other initiatives to level out differences between societies and to eliminate the presumed 'backwardness' of the indigenous Siberians included the distribution of grain and medicine, the cancelling of debts to traders that had accumulated in almost every Siberian family over the centuries, and later obligatory education and attempts to industrialisation (Forsyth 1992: 243, Ziker 1998: 86). While these changes may have provided a significant short-term improvement in the Arctic regions in comparison to previous conditions, it is obvious that in the long run these

⁸ This inequality was maintained between Russians and native Siberians, as is reflected by terms such as 'alien' to denote people in their home country (Slezkine 1994: 84), as well as by Speranskiy's (1822) three-way classification of societies. These were settled, nomadic and wandering people in decreasing hierarchical order and every society had to be allocated to one of these levels (Slezkine 1994: 84, Raeff 1956 (cited by Anderson 2000: 79)).

measures had negative side effects. The interference of the Soviet State with the traditional lifestyle of the Siberian hunters, fishermen and herders disrupted their self-sufficiency and increased their dependency on the Russian state. While the intentions may have been different, this development did not in any way contribute to the foreshadowed 'sovereignty' as desired by the Declaration of Rights.

Other consequences of the establishment of Communism were disastrous for the indigenous population from the start. When Dudinka fell to the Red Army at the end of the Civil War in 1920, Russians and indigenous people were pressed into military service, and some groups retreated into the tundra where hunger forced them to slaughter their own reindeer. As a consequence they had to rely on others who owned more reindeer, thus creating an inequality in wealth that was later vigorously attacked by the same people who had generated it, the Communist Party (Ziker 1998: 85, Forsyth 1992: 248).

In 1924 a special committee was established to defend the interests of the small peoples of the north and to protect them from further exploitation. Its official name was the Committee for the Assistance to the Peoples of the Northern Borderlands, or in short the Committee of the North⁹ (Slezkine 1994: 152). On the Taimyr, these plans materialised most clearly in the building of trading stations, the so-called *faktorii*, along the Khatanga Trading Way to bypass exploitation by commercial merchants and local dealers (Slezkine 1994: 166, Ziker 1998: 82). In addition, shops were opened, schools were built where, and due to the lack of educated local people, the language of instruction was predominantly Russian.

The members of the Committee consisted of high party officials, but also included famous ethnographers¹⁰. Although the official mission of the committee was to protect the interests of the northern peoples, according to Slezkine,

the Committee's true and sacred vocation was to assist the small peoples in their difficult climb up the evolutionary ladder. Cultural progress meant getting rid of backwardness, and backwardness, in the very traditional view of the committee members, consisted of dirt, ignorance, alcoholism and the oppression of women. (Slezkine 1994: 155-156)

⁹ As Forsyth has it, its task was to promote "the planned organisation of the small peoples of the North in respect of economic, judicial-administrative and cultural-medical matters" (Forsyth 1992: 245).

¹⁰ These included e.g. V.G. Bogoraz and L.Ya. Shternberg (Slezkine 1994)

This may explain why in 1926 a law was adopted by which all 'primitive customs' were outlawed, including clan vengeance, bride-price and marriages between minors (Ziker 1998: 90, Forsyth 1992: 244). While such laws were presented as measures to protect the interest of the native population, it cannot have been pure coincidence that they also exactly fitted the Party's idea of how to transform the native Siberians as quickly as possible into workers for the socialist state.

In accordance with the Leninist idea of self-determination and autonomy, which by Stalin was explicitly equated with the eternal fight against backwardness and in favour of progress, autonomous regions were created for the native populations of Siberia, one of them being the Taimyr Autonomous Region. Ideally, every territory should coincide with one nationality and one language. This concern transpires clearly from the ethnographic literature of that time, in particular in situations where this match was not so obvious, as for example for the Dolgans (Anderson 2000: 74). Since from the very start the Dolgans had been described as 'Yakut with Tungus influence' or as a 'mixed people' (Middendorff 1875: 1476), how to classify them with respect to ethnic identity and territory was not obvious. No wonder that we see an increase in ethnographic accounts of the Dolgans in these years trying to clarify this issue (see Section 2.4.2 below). It also explains, in addition to the motivation of the 'fight against backwardness', the intensified attempts of the State to rule out nomadism, which naturally pays no attention to administrative boundaries. In this political climate, the Taimyr (Dolgano-Nenetskiy) Autonomous Region was established on the 10th of December 1930, reflecting the names of the two numerically largest ethnic groups that inhabited the territory (i.e. the Dolgans and the Nenets).

The period of relative freedom and humane changes that had characterised the 1920's came to an abrupt end in 1929 when Stalin started the collectivisation program, which was meant "to exorcise backwardness through a total class war" (Slezkine 2006: 187) and was in his opinion the only real way to progress and to the ideal classless society. However, what to do with societies that have no obvious classes, particularly if they are the most backward societies where progress is most needed? The answer was simple: if there are no classes to battle against, you create them. Previously classless reindeer-herding communities were forced through a stage of an artificially imposed class system, which had to be purged before they were reborn in the ideal society where everybody was equal. Instead of letting them retain their classless social structures, they were forcefully moulded into Stalinist ideology.

Reindeer herders with more reindeer than others overnight became *kulaks*¹¹, shamans and princes became exploiters. Their property could be confiscated for the State, and exploiters themselves were put to work or liquidated. The expropriation of large numbers of reindeer served the State from two sides. On the one hand, it weeded out the exploiting *kulaks* from society, and on the other hand, the confiscation of the reindeer served well the utopian idea to turn the Russian Arctic into the largest reindeer farm (*olen'sovkhoz*) on the planet. Soon the Taimyr would be an enormous reindeer laboratory occupying the surface of Great Britain and containing 20,000 reindeer (Anderson 2000: 49).

Needless to say, such measures were not warmly welcomed by the indigenous population. While there is little documented evidence of armed resistance against the Russians on the Taimyr, the Volochanka rebellion of 1932 showed that it certainly happened. That year, the inhabitants of the Avam tundra received the message that four thousand reindeer were to be expropriated. Horrified by this news, Evenk, Sakha and Dolgan people near the posts of Dolgany, Avam and Volochanka took to arms and killed 20 party members, injured 14 and lost four of their own men. It may not have been a long-term victory but at the time the resolution of the conflict took a surprising turn. Instead of executing the 'rebels' of the tundra, the owner of the reindeer farm was arrested under a charge of theft (Anderson 2000: 49-50). However, such successful opposition was rare, and by the end of the 1930s the majority of the indigenous population belonged to a collective farm, as did 25% of the reindeer in the region (Stetsyuk et al., 1990: 6 in Ziker 1998: 98)).

The black page in history of World War II severely affected the lives of Russians and indigenous people all across the Soviet Union, including the Taimyr. In contrast to World War I, where many indigenous populations were exempt from military service, now nobody was excused, and while the men had to fight for survival on the front line, the women, old men and children struggled for their lives in the villages (Forsyth 1992: 347-350). The War also interrupted the collectivisation process initiated in the 1930's, but it was eagerly resumed after the War had ended. The post-war period is characterised by bringing collectivisation to an even higher level. Many people who had gone through the collectivisation process in the 1930's, had to do this once again in the 1950's in the light of Stalin's

¹¹ *Kulaks* are wealthy and independent farmers, both characteristics, which go against the Soviet idea of a good citizen. They were considered class enemies of the poor peasant, and therefore had to be eliminated.

consolidation (*ukrupnenie*). The collective farms (*kolkhozy*) were fused into even larger state farms (*sovkhozy*) and entire settlements were moved to the *sovkhoz* territory (Forsyth 1992: 362). On the Taimyr this meant that many smaller and unprofitable trading posts or *faktorii* along the Khatanga Trading Way closed down and people were compelled move into more compact areas around the state farms (Ziker 1998: 104).

At these state farms there was no room for individual differences. Thus, the *sovkhozy* became an amalgamation of people of different ethnic origins. Presumably this was inspired by the practical motivations of making the farm function most efficiently, but the mixing of people of different ethnic origins was also part of the plan. Only when people overcame ethnicity and became Soviet citizens instead would the ideal of a completely equal society be realised (Ziker 1998: 106). As a result, clan awareness further disappeared and was replaced by the larger unit of nationality instead. This was not yet quite satisfactory in the grand scheme of Soviet ideology, but it was a step in the right direction. Members of different Evenk clans would now refer to themselves as Evenks instead of naming their clan. It is also the time that the Dolgans, who had so far been described as consisting of different clans, were firmly established as a single nationality (Dolgikh 1963). This kind of development was not unique to the Taimyr. A similar change is testified in Turkic groups where a diversity of clans came to be 'summarised' under the names Khakas and Altai, which are similar situations where several "newly 'consolidated' nationalities occupy compact territories" (Forsyth 1992: 363). What was left of traditional religion disappeared and atheist celebrations, such as the day of the reindeer herder or fisherman took their place (Forsyth 1992: 365).

It is curious that the Dolgans were the only completely collectivised people by World War II, when their reindeer played an important role in the transport of Russians between the Yenisey and Khatanga Rivers (Forsyth 1992: 386). Nonetheless, they were only semi-sedentary. Until the 1970's many families lived in *baloks* (see Section 1.3.1) and tents and they visited the settlements only for supplies or to pick up their children from boarding schools. However, in the 1970's people were forced to permanently settle in proper houses (Ziker 1998:109). Often they had to leave their traditional territory and were planted into villages "for the sake of administrative convenience" (Forsyth 1992: 399), and other smaller settlements were "liquidated" as they were considered non-viable (*ibid.*)

Despite these disruptive measures, material conditions were rather good in the 1970's and 1980's. Frequent flights connected the villages to the towns and allowed many people to go on holidays to the south. There were sufficient consumer goods, which may have had to do with the Dudinka port that in 1978 was opened all year round¹².

2.3.4 POST SOVIET PERIOD (1989 - PRESENT)

While the *perestroika* and the following collapse of the Soviet Union introduced more freedom in certain domains, it caused serious limitations in others, in particular with respect to material goods. This was felt very notably in the remote areas, which through their integration into the Soviet society over the previous 70 years had lost their self-sufficiency and had become dependent on imported goods and services. Through the collapse of the entire system these could no longer be provided. Transport services decreased or disappeared entirely, imported goods became scarce and prices rocketed. With the collapse of the state farms the main provider of employment disappeared, many people lost their jobs and found consolation in alcohol. Too much time had passed to return to the traditional life of hunting, fishing, and reindeer herding in the way that had supported Dolgan families for centuries. The new generation had not acquired these skills very well because there had been no need to do so and also they had different ambitions after having grown up in 'Russian' society.

By now more than 20 years have passed, and although there are some signs of improvement, the situation in most villages still shows many of these features, and people often feel neglected and forgotten by the state. Of course, this is not the whole story. Schools are being run by enthusiastic teachers, club houses organise events and celebrations, but it cannot be denied that living conditions are far from perfect.

¹² Information taken from: www.taimyr24, accessed on January 27, 2012.

2.4 ETHNOGRAPHY

2.4.1 RUSSIAN EMPIRE (1638 - 1917)

From the 17th century onward, the name 'Dolgan' or 'Dulgaan' appears in records kept by Russian tax collectors. However, at that time 'Dolgan' did not yet denote the ethnic group or nationality it represents today, but was used as the name of one particular Tungusic clan. Unfortunately, we do not know which criteria were used for the categorisation of people as Dolgan. If it were based on self-identification of the people, or on the overall package of culture, lifestyle and language, then this Dolgan clan most probably spoke a Tungusic language as well.

However, Ubryatova points out in her discussion of the document from 1638 in which the Dolgan were first mentioned (see Section 2.3.2.2 and Ubryatova 1985: 8) that a match of language and self-identification cannot be taken for granted. The document refers to a headman or prince who had leadership over clans belonging to more than one ethnic group. Considering that power and prestige often play an important role in the choice of language (variety) it is quite possible that the dominance of the ruling group was eventually transmitted to the level of language. Members of the non-ruling group (Evenks) would have learned the language of the ruling group (Sakha), and may have even adopted it in situations of intense contact due to its higher prestige, resulting in language shift. This is important, because there is evidence that such conditions prevailed in the area of the Lena and lower Vilyuy River, where both the Tungusic Dolgan clan and the Turkic Sakha clans were governed by a single Sakha headman (*tojon*), whose name was Dygyna (ibid: 8). Ubryatova argues that this fact may have been an important motivation for the hypothesised language shift in the non-ruling Tungusic Dolgan group to the language of the ruling Sakha, thus providing the basis for the Dolgan language spoken today (ibid: 8, see also Middendorff 1875: 1467, who even mentions a source from 1632).

An attempt to reconstruct the timing of this potential shift is undertaken by the Turcologist Stachowski. In a short paper (Stachowski 1996), in which he refers mainly to Ubryatova's discussion of the 1638 document, he argues that by that year the Dolgan clan must have already shifted to Sakha. However, he bases his argumentation on the assumption that by 1638 the Dolgan clan was already living on the territory of the Taimyr Peninsula, which is, as far as I can tell, a misinterpretation of the facts presented in Ubryatova's work. In his 1996 paper, Stachowski takes two hypotheses as a given: a) between 1628 and 1630 the Dolgans

were still living in the Lena and Vilyuy area, and b) in 1638, they were living on the Taimyr. Evidence for his first assumption comes from historical records written by two Polish convicts. They were sent to the Lena and Vilyuy area to collect *yasak* in these years and it is reported that they encountered the Dolgans in this region. His second assumption is based on the excerpts from the 1638 document cited in Ubryatova, in which he reads that by that time the Dolgans were living on the Taimyr Peninsula. Given these two ‘facts’, he concludes that the migration of the Tungusic Dolgan clan to the northwest must have taken place between 1630 and 1638. Since at that time the majority of the population on the Taimyr Peninsula were Samoyeds (speaking Samoyedic languages), there would have been no reason, perhaps not even a possibility, for the Tungusic Dolgan to shift to Sakha after their arrival on the Taimyr, i.e. after 1638. Thus, he argues, the language shift must have taken place before they started their northward migration and after they had been subsumed under Tungusic populations by the Polish convicts, i.e. between 1628 and 1638 (Stachowski 1996: 129). However, Stachowski’s second assumption is highly questionable. Unless he possesses more detailed information about the 1638 document than is cited in Ubryatova’s work, which he does not seem to, there is no reason to assume that by 1638 the Dolgan had already migrated to the Taimyr Peninsula. The excerpts in Ubryatova’s grammar clearly state that the Dolgan still inhabited the area of the Lena and Vilyuy Rivers, which is a significant distance away from the Taimyr: “And on the Lena river and the mouth of the Vilyuy live Dolgans and Yakuts...” (Russkaya istoricheskaya biblioteka, p. 968, cited in Ubryatova 1985: 8, translation mine).

An additional document reports Sakha and Tungus clans hiding from the *yasak* collectors, and reveals their hiding place by saying that they “lived in the Vilyuy heights and mountains and did not give *yasak* for over two years until 1644.” (Dopolneniya k Aktam istoricheskim, p. 37, cited in Ubryatova 1985: 9 translation mine). After that, Ubryatova continues, the fights between the *yasak*-collecting Cossacks and the indigenous people, including the Dolgans, continued for several decades. Some Dolgans migrated to the east and mixed with the Tungusic Evens, whereas others “lived for a long time in isolation in the heights of the Vilyuy, and then little by little moved to the territory that is the Taimyr Autonomous Region today” (Ubryatova 1985: 9 translation mine). Thus, the 1638 document makes no mention of Dolgan people on the Taimyr, and it does not provide any clue as to whether the shift to Sakha had already taken place or not.

The first explicit statement about the language of the Dolgans appears some 200 years later when the Finnish linguist Castrén and the German naturalist Middendorff were sent on ethnographic expeditions to study the people in the Siberian Arctic and to “clarify a confusing overlap of the peoples inhabiting the lands between the Lena and the Ob’.” (Anderson 2000: 79). By this time, the Dolgans are located on the Taimyr Peninsula. This is also the first time they appear in a context of ethnographic interest, instead of simply as a ‘source’ for *yasak*-extraction (Anderson 2000: 79). The ‘confusing overlap’ concerned the Dolgans in particular, who were sometimes referred to as ‘Tungus’ (their name), and sometimes as ‘Yakut’ (their language). However, both Castrén and Middendorff seem convinced that the Dolgans (or also Dolgasch) are predominantly Sakha, but their identity is mixed with Tungus features as a result of their close vicinity to the latter. According to Castrén, the Dolgans consist of three clans: Dolgan, Edyan and Dongot (Castrén 1856 cited in Middendorff 1875: 1473). While elsewhere these clans are characterised as Tungusic, Castrén identifies them as Sakha (‘Jakutenstämme’), based on his observation that they speak Sakha. Middendorff describes them as “a bunch of emigrated Yakuts” (1875: 1467), but finishes his account with a more nuanced characterisation:

Thus the Dolgan are ... a very distinct, very interesting *mixed people*, in which dominance of the Yakut distinctly emerges in everything.¹³

Both authors characterise the language of the Dolgan people as clearly Turkic (Castrén 1856, Middendorff 1875). Middendorff even describes it as “pure Yakut”, and disagrees with earlier characterisations by Krivoschapkin (1865) who ascribes to the Dolgan people a language similar to Tungusic:

In any case Krivoschapkin is mistaken when he considers the language of the Dolgan to be Tungusic. Without doubt, it is pure Yakut.¹⁴

¹³ “Die Dolganen sind eben ... ein ganz entschiedenes, sehr interessantes *Mischvolk*, bei dem in Allem die Präponderanz des Jakutischen entschieden hervortritt” (Middendorff 1875: 1476, italics and translation mine).

¹⁴ “Es ist jedenfalls ein Irrthum wenn Kriwoshapkin die Sprache der Dolganen für Tungusisch hält; sie ist unzweifelhaft reines jakutisch.” (Middendorff 1875: 1475 translation mine).

2.4.2. SOVIET UNION (1917-1989)

Compared to the sporadic references to the Dolgans during the time of the Russian Empire, they received much more attention from Russian ethnographers and linguists after the establishment of the Soviet Union in 1917. This was not just inspired by an increased interest in ethnography for purely scientific purposes, but was also motivated by political, ideological and administrative changes, which required a clear categorisation of people into 'nationalities', a need which had not existed before.

As the Soviet ideal of being a 'Soviet citizen' gained importance, individual differences were increasingly being wiped out and larger units such as nationalities became more important than the individual clans or tribal affiliations that people used to identify with in the past (see also Section 2.3.3). In addition, the 'administrative clans' that had channelled the collection of tribute during the Russian Empire were reformed by the Soviet government into new administrative units based on nationality to distribute the central state subsidies (Anderson 2000: 82). Therefore it became an ideological as well as a political necessity to divide the population into clear-cut nationalities. Every individual could belong to only one nationality, and terms such as 'Dolgan-Tungus' or 'Dolgano-Yakuts' (Popov, archival data, AMAE 14-1-151, in Anderson 2000: 83) should henceforth belong to the past. These terms already show that the Dolgans occupied an ambiguous position from the start due to the Tungusic as well as Turkic influences in their community. In addition, there was uncertainty with respect to their status as either a single Tungus clan or as a separate ethnic group or even nationality. This explains why the Dolgans have figured prominently in a number of ethnographic and linguistic works between 1917 and 1989, most notably by A.A. Popov, B.O. Dolgikh and E.I. Ubryatova.

All these scholars were indisputably devoted ethnographers, historians or linguists, and there is no doubt that their interest in unravelling the identity of the people inhabiting the fringes of the earth was genuine. However, it is questionable to what extent the published texts correspond to the real opinion of the individual researchers, and to what extent they were edited by Soviet politicians to support and propagate their own convictions. It is nothing new that during the Soviet Period published materials could be severely censored, and there is concrete evidence that certain ethnographic information on the Dolgans underwent the same procedure (Anderson 2000: 82-84), which makes it precarious to rely blindly

on historical materials from that period. Nonetheless, the expeditions of the Soviet ethnographers and linguists provide precious and invaluable information on the indigenous people of Siberia when used with care. A summary of the most influential literature on the Dolgan people, their ethnic composition and their origins will be given below.

In 1930 A.A. Popov set out to the Taimyr National Region (which was established in the same year) to study the ethnography, kinship and material culture of the Dolgans and to “link the Dolgan to [one of] the Turkish, Tungus-Manchurian, or Paleoasiatic groups” (KTsKhIDNI: 28-1-24:1 cited in Anderson 2000: 83). Popov characterises them as the main population of the Taimyr National Region as well as the ‘most advanced’ in terms of culture (see Section 2.3.2.3). With respect to their ethnic composition, Popov distinguishes a ‘core’ or the ‘real Dolgans’, and a group of ‘other Yakuticised people’ who have also become Dolgan. This core consists of four Tungus clans, Dolgan, Edyan, Karanto and Dongot (Popov [1931] 2003: 60), and the other group comprises Russian Tundra Peasants and local Evenks and Sakha people who live in the region.

However, his initially clear definition becomes rather opaque as the description progresses. At present, it is impossible to distinguish the two groups, which evokes the question how Popov himself drew the dividing line in the first place. He concludes by saying that “in fact, the entire native population of the Avam and Khatanga districts can be considered Dolgan” (Popov [1931] 2003: 60) with the exception of the Samoyedic Nganasan and certain Evenk groups south of the Kheta River, thus linking them to a territory rather than defining them by ethnic affiliation. However, despite the confusion, the overall flavour of Popov’s work is a presentation of the Dolgans as a clearly defined nationality with a distinct ethnonym, inhabiting a distinct administrative territory.

With regard to their language, Popov postulates that the Dolgans speak a dialect of Sakha, characterised by a high proportion of Evenki words. This dialect developed in the 18th and 19th centuries when the Tungusic clans moved from the Lena and Vilyuy rivers to the territory of the Taimyr and adopted the Sakha language. Through a common culture and mixed marriages the ethnic boundaries between these groups became less and less pronounced and eventually disappeared, resulting in the people we call Dolgans today.

Anderson, who reviews Popov’s work in detail, remarks that Popov’s unpublished manuscripts differ significantly from the final published version. In his drafts, Popov avoids any explicit statement with respect to the clear-cut

definition of the ethnic composition of the Dolgans and prefers to stick to hyphenated ethnonyms such as 'Dolgano-Yakut', because "the Dolgan don't have a general name of their nationality, every clan has its own name." (Popov [1931] 2003: 13) Not surprisingly, such passages were heavily criticised by reviewers for reasons alluded to above, and had to be rewritten until the Dolgans appeared as the unambiguous nationality desired by the Soviet system. Thus, it is clear that the reviewer's ideological framework penetrated Popov's writing, and that the publicly accessible version of his work does not exactly match Popov's original impressions, to say the least.

The most authoritative material on the identity and ethnic origins of the Dolgan is the work by B.O. Dolgikh. On three expeditions to Arctic Siberia he collected very detailed information on the populations of the Taimyr Peninsula and neighbouring regions. On the first two expeditions, which took place in 1926-1927 and 1934-1935, he went along as a census taker and collectivisation economist, and only on the last one in 1938-1939 was he officially appointed as an ethnographer (Anderson 2000: 85). Dolgikh published a number of studies on this subject, but his most famous work is without doubt 'The origin of the Dolgans'¹⁵, which was published in 1963. In this study he describes the ethnic affiliation, self-identification, and origins of the Dolgan people in meticulous detail, tracing back clans, and sometimes even single individuals, to when and where they were first registered, and how they arrived in their current territory.

In this work, Dolgikh presents the Dolgans as a stable consolidated ethnic group in a similar way to Popov's official version several decades earlier. They are linked to the territory of the Taimyr Peninsula and are clearly separated from the neighbouring Evenk, Sakha and Nganasan populations. However, in earlier work he was not always so certain about the definition of this group, or sometimes even about its very existence. Since unfortunately not all of Dolgikh's original materials are at my disposal, in the following I will rely mainly on Anderson's review of them (Anderson 2000: 74-96).

Dolgikh's first appearance on the stage of the discussion on Dolgan identity is much earlier than 1963. In 1929 he publishes a field report on the basis of his first expedition to the Taimyr as a census taker. In this report he suggests that the patchwork of different peoples on the Peninsula be divided into five 'socio-economical groups':

¹⁵ Original: "Происхождение долган".

- a) Samoyeds-Tavgij
- b) Dolgan (Yakut of the tundra)
- c) Tungus
- d) Yakut (Yakut of the forest)
- e) riverbank Samoyed

These groupings are clearly not based on nationality, but rather on the basis of self-identification and geographical environment, which in turn determines their economic position. This is most clearly exemplified by his categories ‘Yakut’ and ‘Dolgan’: they are both classified as subgroups of ‘Yakuts’, distinguished only by their geographical location (forest vs. tundra) and thus by economic occupation. This is roughly in line with Middendorff’s and Castrén’s identification of the Dolgans as a Sakha tribe. It is also worth noting that this classification was proposed before the establishment of the Taimyr National District as a political unit in 1930 and therefore before the need to create a neat match between the names of the political entity and its inhabitants.

Only a few years later Dolgikh revised his opinion significantly. In 1935 he sent a report to the Provincial Party officials, in which he says that the best classification of the Dolgans has now become ‘Yakutised Evenkis’ (TsGARF A310-18-67: 97-98 cited in Anderson 2000: 86). However, such detailed division of populations would impede cultural-educational work so therefore “it seems possible to consolidate the Yakut, the ‘Dolgan’, and the Tundra Peasants into one national group: the Yakuts” (ibid.). It may seem odd to group the Dolgans with the Sakha after having just classified them as Evenks, but Dolgikh’s decision seems to be founded mainly on the common language among the groups, which was Sakha. After his advice was adopted by the Party officials, the number of ‘Dolgan’ on the Taimyr dropped to zero overnight, the ethnonym disappeared from all official documents, and was replaced by ‘modern’ terms such as ‘Sakha’¹⁶. This is remarkable, considering the fact that only four years earlier a huge administrative territory was established (Taimyrskiy Dolgano-Nenetskiy Natsional’nyy Okrug) carrying the name ‘Dolgan’ to reflect it being the largest population in the region.

Their reappearance occurred as promptly as their disappearance. While there is evidence that Dolgikh himself had been pushing for the return of the Dolgans onto the ethnographic map as early as 1954 (Anderson 2000: 86), this recommendation was recognised only in 1961, after the return of an ethnographic

¹⁶ In these years, other terms of ‘imperial chauvinism’ such as Tungus and Samoyed were also being replaced, with Evenk and Nenets used instead (Anderson 2000: 86).

expedition to the Taimyr Peninsula, which had set out to celebrate the thirtieth anniversary of the Taimyr National District in 1960. On arrival, the members of the expedition were surprised to discover that one of the peoples that determined the name of the district was missing on its own territory. The change of their ethnonym from Dolgan to Sakha had had the consequence that in the entire Taimyr (Dolgano-Nenets) National District there was no official sign of the Dolgans. On return, the head of the expedition pointed out that “there is a complete lack of data on the leading national group of the Taimyr (Dolgano-Nenets) District - the Dolgans”. He continues that those ‘Dolgan’ call themselves ‘Sakha’, but do not identify with the “Yakut character of this term”, by which he refers to the Sakha living across the border in the Sakha Republic. The conclusion was that the term Dolgan needed to be reintroduced, in order to restore the match between administrative and ethnographic boundaries, and to do justice to an apparent difference between the ‘Sakha’ of the Taimyr and the ‘Sakha’ of the Sakha Republic.

Dolgikh was again the right man to do this. His ‘The origin of the Dolgans’ is a confirmation, almost a plea, to recognise the Dolgans as a distinct nationality. If in the past there was uncertainty regarding this matter, so he says, this can be justified by the fact that the Dolgans are a very young nationality, which was still in the process of formation. Today, however, this process is completed and the Dolgans are firmly established and distinct from all their neighbouring ethnic groups. Dolgikh certainly recognises, and even highlights, the ethnic diversity within the Dolgan population, but this does not inconvenience him at all. His *a priori* conviction about their current unity is so strong, that the diverse origins are at most a matter of interest, not a reason to question the appropriateness of merging them into a single ethnic group.

Through exhaustive study of archival materials, Dolgikh (and following him Ubryatova 1985) breaks down the Dolgan population into as many as nine different ethnic components: Dulgan, Dongot, Edyan, Karanto, Yakut, tundra Yakut, Tundra Peasants, Evenks, and Enets. The first three (Dulgan, Dongot and Edyan) he groups together as ‘Dolgan’, the members of the Karanto clan as ‘Evenk’. This differs from Popov’s description, who lists all four of them as Evenks. The Yakuts he specifies as coming from Lake Yessej and the Kotuy and Popigay rivers, and the Tundra Peasants are classified as Russians. The Evenks come from the councils Letneye and Ilimpeyskoye, and the Enets are a few individuals who adopted the “Dolgan dialect of the Yakut language” (Dolgikh 1963: 93).

Using the census numbers from 1926-1927, in the collection of which he participated himself, Dolgikh calculates the proportion with which each ethnic component is represented in the Dolgan population. These numbers are presented in Table 2.2. He comments that these percentages are not based on accurate numbers but that they appear to him as ‘most probable’ (Dolgikh 1963: 128). As will be shown in Section 2.6.3, his estimates are strikingly similar to the latest results from genetic analyses.

Table 2.2. Proportions of different ethnic components in the Dolgan population

<i>Ethnic group</i>	<i>Clan</i>	<i>Percentage</i>
Tungus	Dulgan	50-52%
	Dongot	
	Edyan	
	Karanto	
	Evenk	
Yakut	Yakut	30-33%
	Tundra Yakut	
Russian	Tundra Peasant	15%
Samoyed	Enets	3-4%

A crucial role in the consolidation of these different groups into one “uniform mass of Dolgans” (Dolgikh 1963: 96) is ascribed to the Khatanga Trading Way. This corridor from Dudinka in the west across the Taimyr Peninsula to the east enabled the flow of goods and people, and required more interethnic communication than in other more isolated parts of the Taimyr Peninsula. The trade along the Khatanga Trading Way was as lucrative as it was harsh. Indigenous people had a greater chance to acquire imported goods such as tea, flour, sugar and tobacco, while at the same time they ran the risk to be exploited for their services, in particular for providing transport for the Russian trading caravans, which was a major disruption to the lives of the indigenous population.

Despite these risks, history shows that the Khatanga Trading Way kept attracting people from various ethnic origins, in particular Sakha, different Tungusic groups and Russians. Gradually, intergroup differences became less distinct, and a common mode of subsistence (trade, hunting, reindeer herding and fishing), a common language (Sakha, which served as a *lingua franca*), the adoption of the orthodox religion and intermarriage between the groups (see

Section 2.3.2.3) increasingly obscured the dividing lines between the different ethnic groups (Dolgikh 1963: 136). This blurring and eventual eradication of ethnic boundaries is what Dolgikh describes as 'Dolganisation'. Dolgikh's description pictures this development almost like a chemical reaction, which took place to whoever entered the 'reactor' of the Trading Way. Fuelled by the attraction of goods, services and information, this reactor fused into a unified mass of Dolgans whoever came into its sphere.

While the Khatanga Trading Way may have been a point of interethnic encounters since the 17th century, Dolgikh is convinced that at that time the ethnic identity of the Dolgans as we know it today had not developed yet. In his view 'Dolganisation' started only in the 19th century, when 'a new [name] came into use, [which] testifies that here began forming a new ethnographic community, which did not suit any of the old ethnic names...' (Dolgikh 1963: 107). The establishment of proper trading stations along the Khatanga Trading Way in the 1920's intensified this development and by 1926 the consolidation of the Dolgan as a nationality had in principle been completed (ibid: 106, 137). By that time, he argues, there were almost no families in the area along the Khatanga Trading Way that did not have mixed marriages (ibid: 136) and did not share the Dolgan dialect of Sakha.

The next leap forward in consolidation was the creation of the Taimyr national district, which officially carried the name of the Dolgans. This was followed by the introduction of collective farms and of boarding schools, where people from all different ethnic backgrounds came together and ethnic boundaries were of no importance (1963: 137).

In the light of the great detail with which Dolgikh traces back the component groups of the Dolgans, it is remarkable how easily he sweeps under the carpet the mismatch between their official naming and their self-identification. He admits that most of the component groups do not call themselves 'Dolgan', but refer to their clan names such as Dongot, Edyan, Karanto. The people he classifies as 'Yakut' in fact call themselves 'Sakha' and the Tundra Peasants call themselves 'Yakut' or 'peasant' (ibid: 104-105). However, Dolgikh smoothes over this mismatch with the rather paternalistic explanation that the merging process was a fact, but had not yet been recognised by the people themselves, or as Anderson words it, it was just an "empirical anomaly which only establishes that the Dolgan are a nation in the process of creation" (Anderson 2000: 87). Dolgikh treats the confusing nomenclature of the Dolgans with the same superficiality. Just as he

presents their consolidation into a nationality as a given, he also presents their appearance and disappearance throughout history as simple facts. The complicated ethnic composition of the Dolgan people, and the infelicitous choice of the label ‘Sakha’ for two populations that for Dolgikh are clearly distinct in both ethnic affiliation and language (i.e. the Sakha and the Dolgan), should according to him suffice as an explanation for their impermanent existence throughout history (Dolgikh 1963: 106).

It is striking to observe how each of these key events in the formation of the Dolgan as a separate ethnolinguistic group reserves a prominent role for external factors. Each change of name or identity was brought about by Russian officials and the indigenous people themselves seemed to have no say in these decisions. The literature suggests that the re-establishment of the match between the administrative and ethnographic boundaries of the Dolgans was the mirror image of how the administrative region had been created. Instead of naming the region after the people who already inhabited it, now the people were named after the region they happened to inhabit, and their language seemed to play an equally important role: “Thus, we consider Dolgans the entire current Yakut-speaking population of the Taimyr National Region”¹⁷.

Thus by the end of the Soviet Period, the status of the Dolgans as a nationality had been secured. However, it remains questionable to what extent this happened as a response to the sense of identity of the Dolgan people themselves, or to what extent it was the Russian ethnographers and politicians who created it. Whatever the answer to this question, it does not change anything with respect to the genuineness of their sense of unity today. After all, consolidation may be a matter of a long shared history as much as it may be a conscious decision. The above discussion is only meant to point out the complexity of factors that play a role in such processes.

To give an overview of the diverging opinions on the identity of the Dolgan, a summary of the authors, the clan names as used in the original source, and the associated languages is given in Table 2.3.

¹⁷ “Таким образом, мы относим к долганам все современное якутоязычное население Таймырского национального округа” (Dolgikh 1963: 99, translation mine).

Table 2.3. Different interpretations of ‘Dolgan’ over time

<i>Author</i>	<i>Year</i>	<i>Description</i>	<i>Language</i>
Krivoshapkin	1865	Tungus?	Tungus
Castrén	1860	Sakha tribes, identity mixed with Tungus	Sakha
Middendorff	1875	Sakha tribes, identity mixed with Tungus	Sakha
Popov	1931	4 Tungus clans: Dolgan, Edyan, Karanto, Dongot	Sakha dialect with many Evenk words
Dolgikh	1929	Yakut of the tundra	
Dolgikh	1935	Yakutised Evenks	
Dolgikh	1963	Mix of 9 ethnic groups: Dolgan, Dongot, Edyan, Karanto, Yakut, Tundra Yakut, Tundra Peasants, Evenks, Enets	Sakha dialect
Ubryatova	1985	see Dolgikh 1963	Dolgan
Ziker	1998	Mix of Yakut and Tungus, Tundra Peasant, Samoyed individuals	Creole with Sakha grammar and Evenki lexicon
Anderson	2000	Mix of Sakha, Evenks, Tundra Peasants, Nganasan, Enets	Dolgan

2.5 LANGUAGE

As much as the status of the Dolgan people’s nationality has been a matter of debate, so has been the status of their language. At present it is fully recognised that the Dolgans have their own official language called Dolgan. They have their own spelling system¹⁸, an emerging written literature, educational material, a page in the Taimyr newspaper and a radio programme. However, this recognition of Dolgan as a separate language took place only in the 1970s. Before addressing this issue in more detail, a brief summary is provided regarding the language situation on the Taimyr Peninsula over time.

¹⁸ Dolgan orthography was developed in the 1970’s by A.A Barbolina.

2.5.1 LANGUAGES ON THE TAIMYR

As was mentioned in Section 2.3.1, the earliest inhabitants of the Taimyr Peninsula were probably related to the Chukchi or the Yukaghir people, and so most likely spoke languages that were not related to any of the large language families present in Siberia today (i.e. Indo-European, Tungusic, Turkic, Mongolic, Uralic). The languages spoken by these earliest inhabitants are sometimes subsumed under the name Paleosiberian, but this category is as incoherent as it is controversial (Comrie 1981: 10). From the 2nd century CE onwards, waves of Samoyedic populations moved into the area from the west, later followed by Tungusic clans, presumably with their corresponding Samoyedic and Tungusic languages. From the 9th century onwards, when the influx of Samoyedic people intensified this group spread even further. Thus by the time a new migration wave of Turkic and Tungusic people started moving northwestwards from the Lena and Vilyuy Rivers in the second half of the 17th century, the dominant languages on the Taimyr Peninsula were primarily Samoyedic (i.e. Nganasan, Nenets, Enets) and Tungusic (Evenki). Although Russian fur hunters and tax collectors were also present on the Taimyr at that time, until the 20th century their linguistic influence was insignificant, since most of them did not live there permanently, and they were hugely outnumbered by the indigenous populations (Stern 2009: 388).

With the influx of Turkic-speaking people the balance of languages changed once again, and Sakha became dominant in the region. For this time period, the ancestor language of Dolgan, which I will refer to as Sakha/Dolgan, is often characterised as *lingua franca*, and as a shortcut I will adopt this term as well. However, it needs to be kept in mind that this ancestor of the Dolgan language was more than just a means for interethnic communication. People who joined the open community along the Khatanga Trading Way used lifestyle as well as language as markers of membership in the newly developing social entity, which later identified as Dolgan. This may also explain why Sakha was adopted in domestic spheres by people of different ethnic backgrounds, eventually leading to language shift, rather than remaining confined to trading situations.

This idea is supported by the fact that besides Sakha/Dolgan, there was another language of intergroup communication, called Taimyr Pidgin. In contrast to Sakha/Dolgan, this language *did* remain restricted to trading situations and was never adopted as a first language. Taimyr Pidgin is a Russian-based pidgin heavily influenced by Sakha, which developed from the 18th century onwards, and in

which the ancestors of the present Dolgans are assumed to have played an important role (Stern 2005: 291). According to Stern it was used as a communication system parallel to the standard variety of Russian.

...up to the 20th century two clearly identifiable varieties of Russian were in use on Taimyr, the first being an ingroup variety of the bi- or trilingual group of the Zatundra peasants within the larger community of semi-sedentary newcomers (i.e. the Dolgans), and the second being a pidgin as outgroup variety, which was mainly used to enable communication across the major social divide of the peninsula, namely between the self-segregating Nganasans and the ethnically heterogeneous population of the Chatangskij trakt (i.e. Khatanga Trading Way, E.S.). (Stern 2009: 392)

Taimyr Pidgin was mainly used for communication between the traders along the Khatanga Way and the more seclusive groups of Nganasan people who did not participate in the new community, but only visited the settlements for barter (Stern 2009: 391-392). Now if Sakha/Dolgan only served the purpose of interethnic communication, it is hard to understand why it was not used in the interaction with the Nganasan as well. The identificational value of Sakha/Dolgan with the community along the Trading Way and its function to flag group membership provides an explanation. Nowadays nearly everybody has native command of Russian, and the pidgin is spoken only by a few, mainly Nganasan, individuals older than 75. In the further discussion Taimyr Pidgin will not be treated in detail due to the marginal role it seems to have played in the development of the Dolgan community. However, the fact that Taimyr Pidgin was promoted mainly by the ancestors of the peoples who call themselves Dolgan today (including the Russian Tundra Peasants), shows that Russian-Sakha bilingualism has existed from the early stages of contact with the Russians. This may have had its repercussions not only on the shape of Taimyr Pidgin Russian, but also on the development of Dolgan itself.

Summarising one could say that along with the indigenous Siberian languages and Russian, two lingua francas of quite a different nature were spoken on the Taimyr. One of them served merely the practical purpose of intergroup communication (Taimyr Pidgin), whereas the other (Dolgan/Sakha) had the additional identificational function of binding people together in a new socio-economic community.

The exact motivations for why Sakha/Dolgan occupied this role and not for example, Evenki, remain hazy due to the lack of socio-historical information from that time. However, it is plausible that the relatively large number of Sakha/Dolgan speakers in combination with their alleged prestigious status facilitated the adoption of Sakha/Dolgan as a *lingua franca*. It is interesting to note that during the 17th and 18th centuries the Sakha were expanding not only northwards into the Taimyr, but from the Lena River they spread in all directions, and in many cases their language came to dominate in the new area too (Stern 2009: 391).

Thus, just as the Khatanga Trading Way was an accelerator for the spread of goods and people, the environment of trade, interethnic contacts and intermarriage facilitated the spread and establishment of Sakha/Dolgan in this socio-economic environment. With the increase in interethnic marriages it is plausible that those people who permanently occupied this region (i.e. Sakha, Tungus and Tundra Peasants) began to use the *lingua franca* in private spheres as well, leading eventually to language shift by the non-Sakha groups. This resulted in a variety of Sakha that displays influences from Evenki and Russian, and which nowadays is called Dolgan.

2.5.2 DOLGAN: A DIALECT OR A LANGUAGE?

Over the past three centuries, characterisations of the language variety spoken by the Dolgan have varied from ‘Tungusic’ (Krivoschapkin (1865) in Middendorff 1875) to a dialect of Sakha (e.g. Middendorff 1875, Castrén 1856) and from a ‘Sakha based creole’ (Ziker 1998: 102) to ‘the Dolgan language’ (Ubryatova 1985, Stachowski 1993, Artemyev 2001). This discussion is partly based on linguistic criteria, and partly on the same political and ideological changes that shaped the Dolgan nationality. Even today scholars feel the need to take an explicit stand on the question whether Dolgan is a dialect of Sakha or whether it is a separate language (Stachowski 1993, Artemyev 2001), which indicates that the discussion is still vivid in people’s minds and that the conclusion is not self-evident. The contemporary view is that on the basis of linguistic criteria (e.g. mutual intelligibility), Dolgan may well be considered a dialect of Sakha, but as soon as socio-cultural factors are taken into account, it is clearly a separate language.

From the point of view of language contact studies the classification of Dolgan as a language or a dialect is largely irrelevant. After all, the label of a particular variety as 'language' or 'dialect' does not influence the nature of contact-induced changes or their significance for a people's (pre)history. However, a brief discussion of the different lines of thought is necessary as part of the Dolgan's complex history, as it illustrates how arbitrary and artificial the boundaries are along the continuum of languages and dialects.

Turning a blind eye to the exact details of time and place for the moment, there is common agreement that the ancestors of the present day Dolgans are predominantly Tungus and Sakha groups who migrated northwest from the Lena and Vilyuy Rivers. We have no documented information regarding the languages these individual groups spoke, but it would be intuitive to assume that most of the Tungus clans spoke Tungusic languages (Evenki or Even) and the Turkic groups spoke Sakha. However, as was argued in Sections 1.1 and 2.1, the Dolgan provide evidence that a correspondence between clan and language does not always hold since they have a Tungusic name, but speak a Turkic language. This inconsistency was explained through a scenario of language shift, whereby the Tungusic Dolgan clan adopted the Sakha language, which then spread over a larger area and became the lingua franca for interethnic communication. Supporting evidence for this hypothesis was taken from Ubryatova's reference to the fact that in the 17th century the Sakha and Tungus groups shared a single headman, which may have stimulated Tungusic-Turkic bilingualism in the Tungusic groups, and potentially language shift (see Sections 2.3.2.2 and 2.4.1).

Additional confirmation for an increasingly dominant position of the Sakha and their language is provided by Dolgikh. He notes that by the end of the 17th century in the Olenek region, which used to be occupied by Tungusic clans, 60% of the population had become Sakha (Dolgikh 1963: 114). Dolgikh does not exclude the possibility that the Tungus of this area may already have been bilingual at the time, and he is quite confident that some of them would become so later, in particular the members of the Tungusic Edyan clan, who inhabited this area and are a recurrent component in the description of the Dolgan people in all historical documentation. Dolgikh even goes as far as proposing that the Edyan may have introduced the Dolgan dialect of Sakha to the Taimyr Peninsula (*ibid*: 114).

Despite these indirect historical and demographic facts, conclusions about the languages people spoke at the time remain speculative. The first time explicit mention was made of the language of the Dolgan people, was during the

expeditions of Castrén and Middendorff in 1845-1849 and 1845, respectively. As mentioned above, Middendorff describes their language as ‘pure Yakut’ and refutes earlier identifications of it as Tungus (Krivoshapkin 1865). With the exception of Krivoshapkin, there has been consensus that Dolgan is a Turkic language with certain Tungusic influences, and that it shows a high degree of similarity with Sakha. However, the degree to which either the similarities or the differences with Sakha are highlighted differs strikingly, as can be seen from the range of definitions from ‘a Sakha dialect’ (Middendorff 1875), to ‘a Sakha based creole’ (Ziker 1998) or ‘a separate language on purely linguistic grounds’ (Ubryatova 1985).

It is nothing new that the classification of language varieties as languages and dialects is in fact a continuum (e.g. Ross 2003: 177) and that linguistic criteria such as mutual intelligibility are not necessarily a reliable measure to make this distinction. There are many examples where mutually intelligible language varieties have been granted the status of ‘language’ (e.g. Serbian and Croatian), while very different varieties are considered dialects (e.g. varieties of Chinese and of Khanty). In such cases, the degree of difference or similarity accorded to the varieties seems to be based on political motivations rather than on inherent linguistic properties. After all, the recognition of an official emblematic language fosters a sense of unity, which is important for the establishment of any political unit. Therefore linguistic differences within political boundaries are often glossed over, whereas they tend to be highlighted across political boundaries to underline ‘foreignness’ of the people on the other side of the fence. This point of view is well summarised in Weinreich’s famous quote that “a language is a dialect with an army and a navy”¹⁹ (Weinreich (1945: 13).

Although the Dolgans do not have their own army and navy, the oscillation of the status of their language variety between ‘dialect’ and ‘language’ over time is a good example of the fluid boundary between those two categories, and of the important role politics play in this classification. After the establishment of the Taimyr National Region in 1930, it was important to fuel the growing, partly externally imposed, sense of unity among the people who belonged to this unit. Such processes are speeded up when outsiders, especially scientists, come to study the community in question and ‘objectively’ confirm the commonalities within, and differences across, the groups. This applies to ethnic identity as it does to

¹⁹ Original: פֿלאַך און אַרמיי אַן מיט דײַאַלעקט אַ איז שפּראַך אַ [A shprakh iz a dialekt mit an armey un flot] (Weinreich 1945: 13).

language, and in the case of the Dolgans we see it happening in both domains. What Dolgikh did for the recognition and in a certain sense creation of the Dolgans as a nationality, Ubryatova did in the domain of language. Coincidentally (or maybe not quite), her ‘Language of the Norilsk Dolgans’²⁰, which is the first grammar of the Dolgan language, was published only three years after Dolgikh’s ‘The origin of the Dolgans’. Besides providing a grammatical description of the language, Ubryatova pleads in her introduction for the recognition of Dolgan as a separate language on purely linguistic grounds. She argues that pervasive differences exist between Dolgan and Sakha in the domains of “phonetics, morphology and in particular in the lexicon” (Ubryatova 1985: 17), which according to her could only have formed during a long period of isolated development separate from Sakha, and which suffices to grant it language status on an exclusively linguistic basis.

As alluded to above, the most recent linguistic opinions are critical of this argumentation. In his introduction to ‘The Dolgan language’, Artemyev (2001) stresses the importance of making a distinction between the linguistic criteria and the socio-cultural factors that play a role in the division between dialects and languages, and he finds the linguistic criteria adduced by Ubryatova unconvincing. However, the historical and socio-cultural differences with Sakha are sufficient to classify Dolgan as a separate language (Artemyev 2001: 6), which is supported by Stachowski (1993: 16), when he says that the language-or-dialect-issue is mainly dependent on the “sense of unity of the separate language communities”²¹.

2.6 GENETIC COMPOSITION OF THE DOLGAN

2.6.1 BASIC CONCEPTS

While historical and ethnographic information is essential to understand the history, as well as the present state, of a people, the divergent accounts show that it is not always clear how much credibility should be given to the classification of populations on the basis of archival data alone. Often information on ethnic affiliation was not collected by ethnographers, but by tax collectors whose main concern was of course tax and tribute and not to provide an accurate account of

²⁰ язык норильских долган.

²¹ “Zusammenhoerigkeitsbewusstsein der einzelnen Sprachgemeinschaften” (Stachowski 1993: 16).

the peoples' history. And even if it was collected by ethnographers, we have seen that their goal was not always unambiguous: were their ethnographic accounts intended to describe reality, or to shape reality to fit their politically inspired ethnic classification? Thus, these accounts are not sufficient to disentangle the complex composition of the Dolgan people.

The only way to get a more reliable picture of the ethnic origins of the Dolgans and thus of their prehistory, is by looking at their genetic composition as well. The different proportions of genetic markers, or haplogroups, within the population can give insights into patterns of admixture and migration of the various populations that have resulted in the ethnolinguistic group that carries the name 'Dolgan' today. This section gives an overview of the results of this enterprise, the full account of which is forthcoming (Whitten et al. in preparation).

Genetic markers can be used to study the overall history of populations. Two specific parts of the genome highlight the maternal and the paternal prehistory. For the investigation of the maternal history of a population, it is common to use analysis of the mitochondrial DNA (mtDNA)²². MtDNA is genetic material that is only transmitted in the maternal line and does not, in contrast to autosomal DNA, undergo recombination²³, making it a reliable way to reconstruct genealogies of mutations for the mtDNA (Pakendorf 2007: 330). The paternal history of a population can be studied with the help of analyses of the Y-chromosome, which is only passed on from fathers to their sons. Like mtDNA, most of the Y-chromosome does not undergo recombination and can therefore also be used to trace particular genetic mutations back through time.

Now how can this information be used to study admixture and migration patterns of populations? Important concepts here are the notions of haplogroup

²² A mitochondrion (pl. mitochondria) is a specialised unit in a cell that is involved in a range of processes, an important one being the provision of energy to the cell.

²³ Recombination is a process that occurs during the production of gametes, or reproductive cells. Instead of producing an identical copy of the maternal and a copy of the paternal chromosome, recombination describes the event where part of the maternal chromosome fuses with part of the paternal chromosome during the production of gametes, due to physical overlap of the two chromosomes prior to the cell splitting. In other words, parts of the homologous chromosomes are 'recombined'. This kind of cell division results in a new germ cell, parts of which come from the mother and parts of which from the father, rather than coming entirely from one parent. While the good thing is that this leads to a large genetic variation in offspring, the randomness of this recombination makes these chromosomes unsuitable for the determination of a common ancestor.

and haplotype. Haplogroups are defined by mutations on the Y-chromosome and the mtDNA that are assumed to occur only once in human history, and individuals who share the same mutation are said to belong to the same haplogroup. This type of mutation is called a SNP mutation, which is short for single nucleotide polymorphism (Rubisz 2007)²⁴. Since the Y-chromosome and the mtDNA do not recombine, it is possible to reconstruct phylogenetic trees for these molecules and trace them back in time and space. This is useful because people who share a SNP (and thus belong to the same haplogroup) must share a common ancestor at some time in prehistory. To a certain extent, these SNPs bear similarity to the phenomenon of ‘shared innovations’ in historical linguistics, which are used in a similar way to identify a ‘common ancestor’ of two languages, and thus linguistic relatedness (Pakendorf 2007: 332). Since haplogroups occur in groups of related individuals, particular haplogroups have become associated with groups of populations and are conceived of as a genetic marker of these groups. However, haplogroups do not unambiguously correlate with one ethnic group. They frequently occur in more than one population, in which case the SNP mutations alone are not sufficient to determine the origin of the haplogroup.

In many cases this problem can be solved by looking at haplotypes as well (see below for a definition). For this purpose, longer stretches of DNA are compared, rather than just single SNP mutations. This strategy also enables us to uncover more fine-grained variation between individuals that developed after the SNP arose. For haplotype analysis stretches of the DNA are typed that (in contrast to SNPs) change quickly and are highly variable from individual to individual. For the mtDNA these stretches traditionally correspond to the nucleotide composition of a DNA fragment that is called the hypervariable region (or HVR), but nowadays the entire mtDNA genome can be sequenced for this purpose (see Whitten et al. in preparation). For the Y-chromosome the stretches typically correspond to little chunks of DNA that vary in their copy number (or repeats) and that are called short tandem repeats (or STRs). The set of states for an individual at a given number of loci on the mtDNA or the Y-chromosome is called a haplotype. For the mtDNA the set of states is defined as a particular *sequence of base pairs* at a certain locus on the chromosome; for the Y-chromosome it is defined by the *number of repetitions* of base pair sequences. To illustrate how a difference in STRs on the Y-

²⁴ In fact, in addition to SNP mutations, insertions or deletions of DNA can also define haplogroups. The overarching name for haplogroup-defining mutations is UEP (Unique Event Polymorphism). However, for the purpose of this thesis only SNP mutations are of relevance.

chromosome between two individuals can be determined: individual A has five repeats of the base pair sequence AACT on locus P, and three repeats of sequence TGGC on locus Q. Individual B may have the same number of repeats of AACT and TGGC on these loci on the Y-chromosome, in which case they are said to share the same haplotype²⁵. However, individual B may also differ from individual A and have, for instance, only four repeats of AACT on locus P (instead of five in individual A) and three repeats of TGGC on locus Q (as has individual A). The difference in repeats (which is only one for the current example) defines the genetic distance between the two individuals for this particular locus on the chromosome. Identical base pair sequences at the loci of interest, and thus a shared haplotype, in two individuals is evidence of relatively recent shared ancestry: since haplotypes are established through comparison of quickly mutating regions on the DNA, it is unlikely that they remain unchanged for many generations. On the other hand, large differences in haplotypes within a haplogroup may point to very ancient common ancestry. Hence, haplotype analysis can help identify whether two individuals belong to the same haplogroup through inheritance from a prehistoric common ancestor (in which case haplotypes are unlikely to be shared) or through more recent admixture (in which case they can be shared). In summary, we can say that shared haplogroups, defined by shared SNPs, signify a common ancestor very far back in history, whereas additional shared haplotypes, defined by similarities of base pair sequences (on HVR loci) or number of repeats (on HVR or STR loci), can disambiguate the origin of the haplogroup and distinguish between very ancient and more recent shared ancestry.

Since certain shared mutations, and thus haplogroups, have become associated with groups of populations they can be used to set up hypotheses about possible patterns of inheritance or population admixture in the past. However, the difference between these two scenarios is not easy to establish. Before turning to the results of the mtDNA and Y-chromosome analysis of the Dolgans, it might be useful to briefly mention more generally some genetic outcomes and their associated interpretations.

For the mtDNA as well as the Y-chromosome, a low diversity of haplogroups in a population can be indicative of small isolated populations with endogamy (and resulting genetic drift), while high diversity can be indicative of large

²⁵ In reality, one would include at least five loci, but since this example only aims at an explanation of the principle, only two loci are compared for the sake of clarity.

population sizes and/or admixture. In other words, a low diversity of mtDNA haplogroups can be the result of migration in a scenario where a small proportion of women migrate to a new location and spread only this genetic information in the new community, or of endogamy, when genes are exchanged within the same ethnic group. Admixture with other populations may over time lead to a higher haplogroup diversity. Close genetic distances between mtDNA sequences can be the result of either common ancestry or of admixture, and it is impossible to distinguish between these two scenarios on the basis of mtDNA analysis alone. With respect to the Y-chromosome it is worth noting that a large genetic difference between populations is associated with patrilocality, i.e. a social structure where after marriage the married couple stays in the same location as the husband's parents. This implies that the men stay in the same community, while the women move to different locations, leading to mixing of the mtDNA gene pools, but separation of the Y-chromosomes.

2.6.2. MTDNA ANALYSIS

Analysis of complete mtDNA genome sequences shows that the Dolgan population is, in the maternal line, very closely related to a population that in this study is identified as Yakut-speaking Evenks as well as to the Taimyr Evenks (Whitten et al. in preparation). In this section I will refer to the populations as they are labeled in the genetic study, whereby it is important to keep in mind that the label Yakut corresponds to what I normally call Sakha. The first group, the Yakut-speaking Evenks, lives in the Olenek area and speak, as the name suggests, Yakut (or Sakha). However, they self-identify as Evenks, despite the fact that they do not speak the Evenki language. The second population, the Taimyr Evenks, are a group of Evenks who live on the Taimyr Peninsula. An overview of the populations that are compared in the study, their geographical location and their labels is provided in Map 4.

An analysis of shared mtDNA haplotypes across 21 Siberian populations, including Mongolic, Turkic, Samoyedic, Tungusic and Yukaghir populations, reveals that the highest percentage of shared haplotypes occurs between the Dolgan, the Yakut-speaking Evenks and the Taimyr Evenks, indicating that the genetic distance along the maternal line between these groups is very small. It needs to be mentioned that the mtDNA haplotypes are widely shared across

Siberian populations, which may point to a shared common ancestral gene pool, or it may reflect a historical scenario in which the women moved widely across Siberia, or a combination of both. Thus the mere fact that the Dolgans show genetic similarity with other ethnic groups is not particularly special. However, what is unique is the high percentage of shared haplotypes between the Dolgans, the Yakut-speaking Evenks and the Taimyr Evenks, when compared other ethnic groups in Siberia.



Map 4: Peoples and locations where genetic samples were collected

More precisely, the Dolgan share 60% of exact mtDNA sequences with the Yakut-speaking Evenks from Olenek and about 48% with the Taimyr Evenks. The Taimyr Evenks and the Yakut-speaking Evenks share in turn about 50% of exact mtDNA sequences with each other. Even between subpopulations, such as for example the Central Yakuts and the northeastern Yakuts, the percentage of shared haplotypes is not as high (about 38%) as between the Dolgan and their geographically adjacent, but ethnolinguistically different, groups. Thus, this picture suggests that there has been contact in the maternal line between Dolgans,

Yakut-speaking Evenks and Taimyr Evenks and that women married into communities that were ethnolinguistically different from their own. As may be remembered from Section 2.3.2.3, this is fully commensurable with the table of marriages that was provided by Dolgikh.

2.6.3. Y-CHROMOSOME ANALYSIS

While in the mtDNA the Siberian populations share a lot of their genetic material, the Y-chromosome shows more differentiation across populations. As mentioned above, this could be indicative of patrilocality, which matches the ethnographic descriptions of marriage patterns of both Turkic and Tungusic populations.

Analysis of the Y-chromosome in a number of Siberian populations shows that certain haplogroups, referred to arbitrarily by letters of the alphabet, are strongly represented within certain ethnic groups. The codes of some haplogroups found in the Taimyr populations, and the ethnic group with which they are associated are shown in Table 2.4.

Table 2.4: Haplogroups and their associated populations

HAPLOGROUP	ETHNIC GROUP
C	Northern Tungusic (Evenk, Even), Mongolic
N2	Samoyedic, Tungusic
N3	Yakut, but also Uralic and other northern Eurasian populations
R	European

As can be seen from the table, haplogroup C is associated with northern Tungusic populations, N3 is a marker of Yakut as well as of Buryats and Uralic populations all the way to the Finns. Despite this ambiguity, N3 has been identified particularly as a Yakut marker, since 94% of the Yakut men carry it in their genome (Pakendorf et al. 2006). Moreover, haplotype identification through STR analysis has shown that the STR haplotypes in these Yakut men show a high degree of similarity, so we can confidently say that within the men that were sampled for this study, haplogroup N3 is a marker of shared Yakut ancestry. N2 is generally found in high frequency in Northern Samoyedic populations (44.9% in Forest Nenets, 74.6% in Tundra Nenets, 92.1% in Nganasan, 77.8% in Enets, but here the sample size is only

9; Karafet et al. 2002) as well as Tungusic populations (in addition to the numbers in the table below, the Central Evens (from Topolnoe) have 37.5% (Pakendorf et al. 2007)). Finally, haplogroup R is associated with Europeans. Now the representation of those haplogroups within a selection of the Taimyr populations and relevant groups for comparison looks as follows:

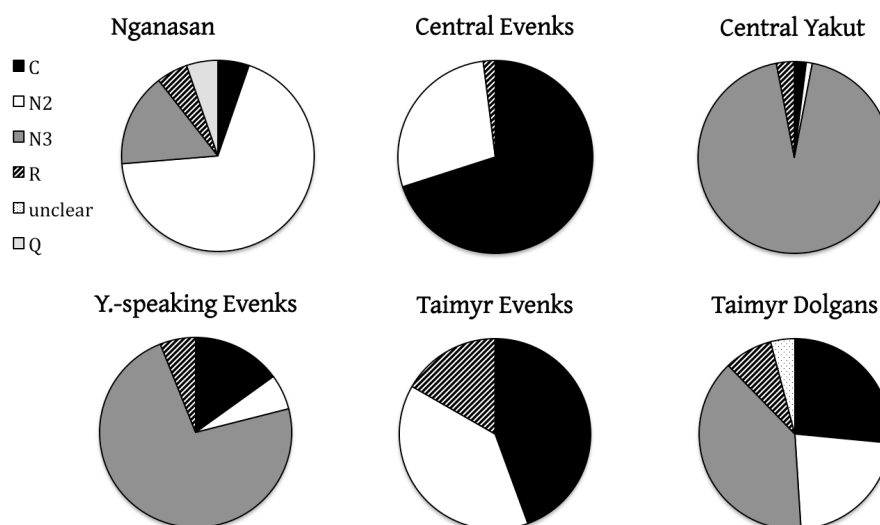


Figure 2.1: Complements of Y-chromosomal haplogroups in north Siberian populations

As can be seen from figure 2.1 in most populations one haplogroup is dominant: in the Nganasans it is N2, in the Central Evens it is C, and in the Central Yakuts and the Yakut-speaking Evenks it is the Yakut marker N3. Compared to this picture, the Taimyr Evenks and in particular the Dolgans demonstrate a more diverse profile. In the Taimyr Evenks haplogroups C and N2 are present in almost equal proportions, and in the Taimyr Dolgans all three markers (C, N2 and N3) are present in comparable frequency, none of them being evidently dominant. On the basis of haplogroup analysis, it looks like the Dolgans share a common ancestor with the Yakut (N3), the Tungus (C), Samoyeds/Tungus (N2) and Europeans (R). However, as was mentioned earlier, some of the haplogroups are not unequivocal with respect to the ethnic group they are associated with, and in order to be sure about the origins of their haplogroup it is necessary to undertake haplotype analysis as well. For N3, STR haplotype analysis shows that this haplogroup in the Dolgans is shared with the Yakut population, as expected on the basis of historical

sources. R, which is a haplogroup found in European populations, turns out to be identical to Y-chromosome haplotypes of Russian men, which is evidence of recent gene flow from Russians into the Dolgan population. For N2, which can be associated with Samoyedic as well as with Tungusic populations, haplotype analysis was not able to disambiguate between these two possibilities. The exact haplotypes found in the Dolgans were shared with Evenks and Samoyedic individuals in approximately equal proportions. Thus, the proportion of haplogroup N2 in the Dolgan population can either point to a Samoyedic or a Tungusic common ancestor. In the first case this would result in a diverse profile of Turkic, Tungusic and Samoyedic haplogroups with a slight dominance of the Yakut marker (approximately 40%). However, the second scenario would support a distribution in which Tungusic haplogroups clearly dominate the picture, even more than the Turkic haplogroups N3, despite the fact that they speak a Turkic language. At this point it is worth mentioning that this picture would be strikingly similar to the ethnic composition suggested by Dolgikh in 1963, who based his picture purely on archival materials. The striking similarity between the two charts as they would look if haplogroup N2 is had Tungusic origin is given in figure 2.2.

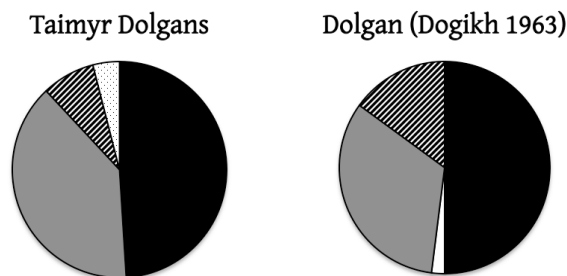


Figure 2.2 Ethnic composition of the Dolgans based on Y-chromosomal haplogroup analysis (L) and on registered marriages (R).

While earlier published data on the Dolgan Y-chromosome display a different distribution of haplogroups, in particular with respect to the frequency of haplogroups associated with European (Russian) ancestry, the dominance of Tungusic haplogroups is confirmed by Karafet et al. (2002). They found the following haplogroup frequencies for the Dolgan population: 37% of the sampled individuals belonged to haplogroup C (typically associated with Tungusic population), 12% to N2 (associated with Tungusic or Samoyedic groups), 22% to N3

(associated with Yakut), and 19% to R and I (associated with European populations). For an easier comparison these proportions are represented in Figure 2.3:

Dolgan (Karafet 2002)

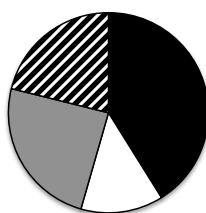


Figure 2.3: Y-chromosomal haplogroups in the Dolgans according to Karafet et al. (2002)

Regardless of the differences, both analyses show a high frequency of the haplogroup associated with Tungusic (Evenk) populations, which reflects that Evenk males must have moved into the Turkic community. Whether these were larger groups of Evenks who moved into the Dolgan community and learned their language, but otherwise remained relatively independent of the Dolgan/Sakha people, or whether the relocation was accompanied by intense intermarriage with individuals from other ethnic backgrounds is impossible to tell from these data. The only fact we can establish is that both mtDNA analyses as well as Y-chromosome analyses give evidence of close contact between the Dolgan/Sakha and Tungusic populations. However, on the basis of Dolgikh's marriage table we can assume that there was a significant amount of interethnic marriage as well.

2.6.4. INTERPRETATION

Both mtDNA and Y-chromosome analyses show close contact between Turkic and Tungusic populations in northern Siberia. What can this information tell us about admixture and patterns of migration, and how does it affect the interpretation of language data?

The mtDNA analysis has shown that the Dolgans, the Yakut-speaking Evenks and the Taimyr Evenks share a high proportion of haplotypes, which means that women were exchanged between these groups. From a genetic point of view, these groups can even be conceived of as a single population, as is indicated by the so-

called F_{st} value. In genetic analysis, the F_{st} value determines genetic distance between populations and is used as a measure of population differentiation. If this value is zero or non-significant, the difference between the populations is so small that it can be conceived of as one unit. For the Dolgans, the Yakut-speaking Evenks and the Taimyr Evenks, the genetic difference was shown to be non-significant (see Whitten et al. in preparation). However, despite the fact that they live in geographically adjacent areas and have for a large part a similar lifestyle, the three groups do not self-identify as one population, and they refer to themselves by different ethnonyms. In addition, there is a linguistic dividing line within the group: while the Dolgans and the Yakut-speaking Evenks speak a Turkic language (Dolgan and Sakha, respectively), the Taimyr Evenks speak a Tungusic language (Evenki). This means that part of the population of Dolgan, Yakut-speaking Evenks and Taimyr Evenks (or at least the women who married into other groups) must have adopted a different language at some time in the past. However, purely from the genetic data nothing can be inferred with respect to the extent to which this happened or about the direction of such a possible shift. On the basis of the mtDNA alone, all three groups could have been Turkic-speaking and the Taimyr Evenks could have shifted to the Tungusic Evenki language. Alternatively they may have been all Tungusic-speaking groups, of which the Dolgans and the Yakut-speaking Evenks adopted the Turkic languages Sakha and Dolgan, and finally, they may have been Turkic-speaking and Tungusic-speaking groups who intermarried.

In the paternal line we have seen that almost all investigated populations expose a certain diversity in haplogroups, but that in each population one of the haplogroups C, N2 or N3 is represented most prominently. An exception to this pattern is the Dolgan population, which shows comparable frequencies of haplogroups C, N2 and N3, indicating that a genetic contribution from Samoyedic/Tungusic, Tungusic and Turkic males is present in the population in almost equal proportions. In theory this could point to a very ancient ancestor that was common to all three populations. However, the haplotype sharing with the Sakha for haplogroup N3, and with the Evenks for haplogroup C that was demonstrated through STR analysis shows that more recent admixture is a more plausible explanation for this diversity. The origin of haplogroup N2, which is associated with Tungusic and Samoyedic populations, could not be determined with certainty.

Of course intermarriage and migration are not the only ways for Y-chromosomal genes to enter a population. It could also happen through events of

rape or one-time physical contact between male and female individuals, but the high extent to which the different haplogroups are found in the Dolgan population makes this scenario highly implausible as a primary explanation. Since the Turkic, as well as the Tungusic, populations are patrilocal, intense marriage of males from different ethnic backgrounds into the community is also unlikely to have happened.

A more plausible explanation for the diversity in haplogroups among the Dolgans is that groups of males from various ethnic backgrounds, and in particular Evenks, moved to the area where the present-day Dolgans live and became part of the new community by adopting a new lifestyle of trading along with reindeer herding and adopting the Sakha/Dolgan language. Whether these males then intermarried with women from other ethnic groups, or whether they rather interacted more with females who came with them cannot be determined on the basis of these data. However, Dolgikh's marriage table shows that interethnic marriages were common and if it is true that the newcomers adopted a new language, it is unlikely that they only interacted with their own people. If they did so, there would have been no need to adopt a different language in the first place, and they probably would not have become integrated completely into the new community.

It also remains unclear on the basis of these data which populations moved into which community, in other words, the direction of admixture. Technically, the distribution of Y-chromosomal haplogroups in the Dolgans could be a reflection of Turkic men moving into Tungusic groups, or vice versa. The fact that the Dolgan speak a Turkic language today may point to Turkic as the dominant language at the time when other populations came into the community, and that the newcomers therefore adopted Sakha/Dolgan. While this interpretation is the most plausible on the basis of historical records, the genetic data alone do not give support of one direction over the other. The historical and ethnographic data that were presented above, in combination with the analysis of contact-induced changes in the language that is still to come, is intended to help find answers to this question.

2.7. SUMMARY

Throughout history we have seen that there has been little consensus on the ethnic composition, moment of formation, or language of the Dolgan people. However, a review of the historical, ethnographical and genetic information conspires towards recognition of the view that the Dolgans are of multiethnic origin, with the main components being Tungus (Evenki), Turkic (Sakha) and Russian. For different reasons these groups moved to the southern Taimyr in the second half of the 17th and beginning of the 18th centuries. Here, initial ethnic boundaries based on descent gradually faded, and they were exchanged and complemented by identity formation on the basis of shared activities (trade), language and ecological zone.

Contact between Turkic and Tungusic groups probably existed as early as the 17th century in the area of the Lena and Vilyuy rivers, but the formation of the Dolgans as a separate ethno-linguistic group took place later. While this process of 'Dolganisation' may have started in the 18th century, their official establishment as a separate ethnic group only took place in the 20th century, under the influence of Russian politicians and state ethnographers, who had no space in their ideological framework for the fluid ethnic boundaries and identity continua that seem to have been present amongst these groups. Most likely, the foundation for today's Dolgan community was formed in the second half of the 17th and the beginning of the 18th centuries when Sakha and Tungusic groups (including the one named Dolgan) moved from the Vilyuy and Lena rivers to the southern Taimyr. Although we have no records of the languages they spoke, it is possible that even back then there was some Tungusic-Turkic bilingualism among the ancestors of the Dolgans, as is suggested by the fact that Tungusic people were ruled by a Sakha headman. This could have involved incipient bilingualism in Sakha in the Tungusic Evenks.

The mutual adaptation of people from different genetic and geographic origins continued after their arrival on the Taimyr in the late 17th and early 18th centuries. The different ethnic groups that engaged in the life of the Khatanga Trading Way adopted trade alongside their traditional activities such as reindeer herding, hunting, and fishing, and grew closer to each other genetically, culturally and linguistically. It will be remembered that the early 19th century was the time of the calamitous attempt to populate the Taimyr, when many Russian peasants arrived in the area around the Khatanga Trading Way and had to adopt the native way of life in order to survive. Thus, although different ethnic groups had arrived

for different reasons, they shared one thing: they were all newcomers to the southern Taimyr, whether Vilyuy Evenk, Lena Sakha or Russian peasant. They were all in a phase of adaptation to a new way of life in a new geographic environment. The dominance of recent immigrants and the absence of any strongly established groups (except the Nganasan who lived further north and barely engaged in the life around the Trading Way) may have made the fading of existing ethnic boundaries a natural phenomenon. Finding a new common unity and identity may have been more essential in the struggle for survival in new inhospitable lands than restricting oneself to the small group of relatives and retaining one's old identity. In this context it also seems natural that this new common identity was based more on shared occupation, ecological zone and language than on descent (Anderson 2000: 91-96). This process of dissolution of ethnic boundaries intensified over the next century or so.

Genetic analyses support the historical and ethnographic accounts. They show that there has been admixture of Sakha and Tungusic groups in the maternal as well as in the paternal line, whereby the similarities in the maternal line are so striking that there must have been a significant amount of marriages of women between the ethnic groups. The fact that the Dolgans nowadays speak a Turkic language implies that in one of the two groups language shift must have taken place. Although from the mtDNA alone we cannot tell the direction of the shift, we know from historical records that Sakha became the *lingua franca*. Therefore we can assume that the Tungusic groups gradually shifted to Turkic Sakha rather than the reverse.

The analyses of the Y-chromosome (i.e. the paternal line) confirms this scenario. The data show that Dolgans are the only group for which the haplogroups C, N2 and N3 are represented in almost equal proportions. Most plausibly, this is indicative of a historical event where men from different ethnic backgrounds moved to the area along the Trading Way, and adopted the *lingua franca* of the area, Sakha. The question of whether these men on arrival only interacted with people from their own community cannot be answered by the genetic results, but the linguistic and ethnographic data give important clues: the adoption of a new language, in particular a *lingua franca*, only seems to make sense when there is a significant amount of interethnic communication. Interaction with other ethnic groups is strikingly confirmed by Dolgikh's data on marriage patterns, which show that only 37.5% of marriages took place between members of the same ethnic group.

The next question to ask is how this complex history is reflected in the language of the Dolgan. There is agreement on the close similarity between Dolgan and Sakha, but if the language shift scenario is true, then we would expect some traces of a Tungusic substrate in the Dolgan that is spoken today. Similarly, if the Dolgans themselves have been bilingual in Russian for some time, this may be noticeable in their current speech as well. It is possible to simply compare standard Sakha and Dolgan and note down the differences. However, in order to attribute meaning to the differences, and to make inferences about what they can tell us about Dolgan prehistory, it is necessary to link the findings to a theoretical framework. Therefore the next chapter will provide an overview of the most relevant ideas from language contact theory, bilingualism and language acquisition. Without pretending to be comprehensive, this background knowledge will equip us with the framework we need to formulate hypotheses about: a) what linguistic changes in Dolgan we might expect; and b) how to interpret the attested changes.

3.1 THE FIELD OF LANGUAGE CONTACT STUDIES

3.1.1 DEFINITION AND BRIEF HISTORY OF THE FIELD

In the most trivial sense, language contact occurs whenever more than one language variety is spoken in the same place at the same time (Thomason 2001: 1). However, such settings are only of interest for the study of contact-induced change when the speakers engage in interaction, in other words, when ‘language contact’ becomes ‘speaker contact’.

While ‘language contact’ has become the standardised label for a subdiscipline of linguistics, it can be misleading since it implies that languages can be in contact independently of their speakers. It evokes a scenario where languages resemble autonomous entities, able to evolve, change and maintain contacts beyond the sphere of their speakers. To eliminate this obvious delusion from the start it is important to remind the reader of a statement with which most linguists would agree, namely that “it is not the languages that innovate, it is the speakers who innovate.” (Milroy & Milroy 1985: 45). Although in this dissertation I will follow common practice and use the term ‘language contact’, it is important to keep in mind that it is meant as a shortcut to refer to contact between speakers and the consequent linguistic variation and change.

Thus the study of language contact is concerned with the investigation of the linguistic consequences of encounters between people who speak different languages or language varieties. Questions to be asked include: which factors shape the linguistic outcome of a contact situation? Are there any linguistic restrictions on the kinds of possible transfer between languages? What is the role of psycholinguistic processing and production mechanisms of the bilingual brain during this process? Is there a correlation between the social setting in which the contact takes place and the linguistic outcome? If there is, is it possible to predict the linguistic outcome from the social setting, and reversely, can we reconstruct past social situations through the analysis of contact-induced linguistic change?

Language contact may be studied at the level of the individual, as in the case of a bilingual person (3.1), as well as at the level of the community in the case of communities where more than one language is spoken (3.2). Here not every individual of the community has to be bilingual, and proficiency in the languages may be distributed across the community to varying degrees.

(3.1) Two or more languages will be said to be IN CONTACT if they are used alternately by the same persons. (Weinreich 1953: 1)

(3.2) Language contact occurs when speakers of different languages interact and their languages influence each other. (Matras 2009, preface)

It follows that the linguistic consequences of encounters between people never occur in a social vacuum and that the importance of the social context in which linguistic changes occur is undeniable. This idea appears at intervals in the literature on language contact (e.g. Weinreich 1953: 3, Thomason and Kaufman 1988: 35, Johanson 2002: 307), and today the significance of extra-linguistic factors in the explanation of contact-induced change is recognised by most scholars (cf. Thomason 2001, 2010, Johanson 2002: 308, Muysken 2010, Ross, 2003, forthcoming).

The first milestone emphasising the multi-faceted nature of contact situations is Weinreich's book *Languages in Contact* (1953). While much of this book reads as a research plan of which many details are still unexplored rather than as the presentation of results, Weinreich argues convincingly that

[a] full account of interference in a language-contact situation, including the diffusion, persistence, and evanescence of a particular interference phenomenon, is possible only if the extra-linguistic factors are considered. (Weinreich 1953: 3).

Being a linguist, and more specifically a linguist influenced by the structuralist theory that was dominant at the time, Weinreich acknowledges the importance of purely structural and typological factors in contact-induced change, but he emphasises that “on an interdisciplinary basis research into language contact achieves increased depth and validity.” (Weinreich 1953: 4). Many of Weinreich’s valuable insights and suggestions for research temporarily faded into the background with the rise of generative linguistics. In this framework, language is an autonomous system characterised by the principles of Universal Grammar, which is innate in the brain of every individual. Since these principles are assumed to be universal to humanity, understanding them in one person would be sufficient to understand the language system in general. Such ideas obviously reduced the significance of comparative linguistics and of extra-linguistic factors in the explanation of language development and change.

Attention to the role of social factors in contact situations was reawakened by the publication of Thomason and Kaufman’s influential book *Language contact, Creolization and Genetic Linguistics* in 1988. In this work they present a theory of language contact, postulating correlations between social contact situations and their linguistic outcomes (see Section 3.1.4.1 for more detail). They revived some of Weinreich’s interdisciplinary ideas, while at the same time taking them to a more advanced level by setting up an explanatory and predictive model. Thomason and Kaufman’s book became one of the most influential and widely cited textbooks on language contact, and can, in a sense, be seen as a trigger for the revitalisation of interest in the field. The publication of this work was followed by an increase in descriptive case studies of contact situations, as well as theoretical models, and the dotted line of sporadic publications on the topic fanned out into a wide diversity of different research programs from the end of the 1980’s onwards. While the elaborated details of these programs are innovative, many hinge upon subparts of the interdisciplinary research sketch that Weinreich advanced some three decades earlier.

This chapter will introduce a small selection of these approaches in a nutshell. Since a complete overview of the field is unfeasible, as well as undesirable for the current purpose, I will discuss the main lines of thought from four approaches that have proven useful for the interpretation of the data in this study. The discussion is organised thematically rather than chronologically, and groups together contributions of scholars with similar theoretical perspectives on

what determines the linguistic outcomes of a contact situation. Obviously, this separation into categories is artificial and there are many points of overlap and intersection between the approaches. However, such a division will be helpful for the sake of presentation.

The discussion begins with a summary of the structuralist approach, according to which purely linguistic factors determine the limits of language change. Subsequent research has shown that language structure is not the only relevant factor and that psycholinguistic and sociolinguistic factors play an equally influential part in the linguistic outcomes of contact situations. Since the ultimate origin of language change is the individual, the obvious starting point is a discussion of the psycholinguistic processes that take place in the bilingual brain. This includes situations of second language acquisition (Section 3.1.3.1) as well as bilingual first language acquisition (Section 3.1.3.2). Equipped with this background knowledge, social dynamics at a community level will be discussed in Section 3.1.4. While an impressive amount of research has been done to describe contact-induced change at the level of both the individual and the community, little is known about the link between them. This topic is briefly touched upon in Section 3.1.5 in the discussion of the role of social networks in the spread and adoption of linguistic innovations in a community, thus also beginning to establish the connection between synchronic variation and diachronic change.

3.1.2 STRUCTURAL CONSTRAINTS

The issue as to what can and cannot be transferred across languages has been approached from various angles. According to the structuralists, the range of possible kinds of transfer is determined by the structural properties of the languages in contact, more specifically by: a) the structural integration of the transferred element; and b) the typological similarity between the languages in contact (e.g. Weinreich 1953, Moravcsik 1978). The structuralist school teaches that cultural elements, including language, are part of a coherent structural system and receive their meaning solely through their relations with other elements within the system. This underlying network of relationships is stable and remains constant despite surface variation. According to the structuralists, this explains why languages are resistant to foreign elements that do not match, or would threaten the integrity of this deep structure. Therefore the assumption with

respect to the first constraint of structural integration is that structurally highly integrated elements, such as bound morphology, are less likely to be transferred than structurally independent elements, such as lexical items.

A convinced advocate of this line of thought is Meillet, who defines language as “une système où tout se tient” [a system where everything is interconnected] (Meillet 1915: 463) and which consists of three components: a phonetic system; a morphological system; and a vocabulary (Meillet 1948: 83). The first two systems are closed, structurally highly integrated and internally more stable, whereas the lexicon is open (but see Thurston 1987, among others, for a different opinion). Therefore, the phonological and morphological systems are more resistant to foreign influence than the lexicon.

Nonetheless Meillet concedes that phonological and morphological contact-induced changes are not entirely impossible, and that sometimes even the stable parts of languages are affected by contact (Meillet 1948: 85). For the phonological system, he adduces the example of the Germanic phoneme /w/ which was introduced into Gallo-Romance in forms like *werra* ‘war’, and the Persian relativizer *ki* which occurs in Turkish. However, for the phonological elements he argues that they are unstable and temporary, since *werra* soon changed into *guerre* ‘war’, replacing the Germanic phoneme /w/ with Romance /g/. With respect to *ki* in Turkish, Meillet claims that its use bears great similarity to a lexical item, and that speakers of the recipient language may conceive of it as such (Meillet 1948: 87). However, he does not elaborate on the criteria on which this judgement is based. Thus, despite such instances, Meillet concludes that:

It follows [...] that one borrows from a foreign language little words with grammatical value; one hardly borrows real grammatical forms. Thus one arrives always at the same conclusion: what is borrowable is essentially vocabulary items.¹

A similar opinion is manifested in the work of Sapir (associated with the school of American structuralism), and Jakobson (an adherent of the Prague Circle) albeit in a more moderate form. Both Sapir and Jakobson propose that it is possible for languages to accept foreign grammatical, i.e. structurally integrated, elements into their system, “provided that these new variations... are in the direction of the

¹ “Il arrive [...] qu’on emprunte à un langage étrangère des petits mots à valeur grammaticale; on n’emprunte guère de formes grammaticales. Ainsi l’on est toujours ramené à la même conclusion: ce qui s’emprunte, ce sont essentiellement des éléments de vocabulaire.” (Meillet 1948: 87).

native drift” (Sapir 1921: 200) and “correspond to the natural tendencies of development” of the recipient language (Jakobson [1938] 1962: 241).

This links to the second constraint that grammatical structures can be transferred only between languages with a similar typological profile. According to this hypothesis, one would for instance not expect languages to change from being head-initial to being head-final, or for agglutinating languages to adopt synthetic elements. However, subsequent research has provided clear counterexamples to this claim (see for example Ross 1996, 2008, Thomason 2010).

Harris and Campbell point out that the implication of the above-mentioned constraints would be that transfer in language contact situations would only be possible from analytic languages to more synthetic languages and not the other way round, since structurally independent elements are more easily transferrable (Harris and Campbell 1995: 31, Van Coetsem 1995: 67, 2000: 31, 111). However, this appears not to be the case. While both constraints are not unmotivated, and may reflect cross-linguistic tendencies, they are anything but absolute (for more details see Harris and Campbell 1995: Chapter 6). The transfer of elements between typologically similar languages is intuitively more likely, and unbound elements may certainly move around more easily and frequently than bound elements, but over the years abundant evidence has accumulated showing that there are too many counterexamples for such linguistic constraints to qualify as a rule.

A classic example is Asia Minor Greek, a dialect of Greek that has been in intense contact with Turkish over a long period of time. In this variety of Greek, phonemes and phonological rules have been adopted from Turkish, resulting in the introduction of new phonemes into its phonological system (ö, ü, ĩ, ĉ, j), and the adoption of vowel harmony, a phenomenon that is not present in other varieties of Greek (Dawkins 1916: 39, 41, Janse 1998: 524). An example of morphological transfer is provided by Resígaro, an Arawakan language spoken in northwest Amazonia. Seifart (2011: 17) shows that this language has undergone substantial influence from Bora, an unrelated Witotoan language spoken in the same region. In Resígaro, not only lexical items, but also a large number of bound morphemes, including quantifiers, pronominal forms and numerals appear to be of Bora origin. The typological profile of Ossetic provides a counterexample to the claim that transfer can only take place between typologically similar systems. Ossetic, which is an Indo-European language, has adopted many features from its Kartvelian neighbour Georgian (Thomason 2010: 42). In addition to a large number of loanwords, these include the introduction of agglutinative morphology into a

system that used to be largely flexional, and a more rigid SOV order. Another example of contact-induced influence among typologically dissimilar languages is the introduction of finite relative clauses in literary Dravidian languages as a result of contact with Indo-Aryan languages (Gumperz and Wilson 1971, Nadkarni 1975, Emenau 1981).

Thus, while structuralist linguistic constraints may reflect cross-linguistic tendencies in the types of transfer that occur, the data show that they do not account for the range of attested phenomena. It seems that contact-induced change is more than a predictable chemical reaction between two linguistic systems, and that for an accurate understanding of its underlying motivations we need a better comprehension of the bilingual human brain, and on a larger scale the bilingual community, in which language is generated.

3.1.3. PSYCHOLINGUISTIC CONSTRAINTS

The relation between language and the brain has been a topic of interest in linguistic research for a long time. As a hearer, we witness and interpret the end product of the speech process, but what actually happens in the speaker's brain remains invisible to the naked eye. However, valuable cues about the psycholinguistic processes in language production and processing are provided by experimental psycholinguistic research, speech error analyses, and by EEG scans that measure brain activity in different regions of the brain.

Since contact-induced change can only take place when at least part of the speech community is proficient to a certain degree in more than one language, research on bilingualism is of particular importance for the understanding of the underlying psycholinguistic mechanisms of this type of change. Bi- or multilingualism is of course not an innate characteristic of an individual, but is acquired during one's lifetime as a child, an adolescent or an adult. This deserves mentioning since the age at which the second language is acquired appears to have significant influence on the level of proficiency that a person can attain in his second language². While individual variation makes it hard to generalise, the

² Although I will be referring to second language acquisition and bilingualism most of the time, in language contact situations multilingualism is certainly no exception either. However, since most of the experimental research has been carried out in bilingual settings, we can only talk with any

tendency is that the earlier in life a language is acquired, the more native-like proficiency can be. This thought is formulated more specifically in the so-called critical period hypothesis (Lenneberg 1967: 179). According to this hypothesis there is a qualitative difference in the way people learn a language before and after this period, which is reached some time during puberty. After this point, it is thought that children are unlikely to acquire a language on a native level of proficiency. It is thus necessary to distinguish between two kinds of acquisition related to different age groups. First, there is second language acquisition (or L2 acquisition), which applies to adolescents and adults who learn their second language after the critical period. Depending on aptitude and attitude, they can still reach very high levels of proficiency, but it is very rare to find individuals in this group who master their second language with native-like proficiency and fluency. The second kind applies to children who grow up bilingually, and learn both languages natively. Since for these children both languages are their mother tongue, the distinction between a first and second language becomes irrelevant, and a more appropriate label for their language learning process is bilingual acquisition.

Over time the reputation of bilingualism and its presumed effects on the human mind have changed significantly. While during the first half of the 20th century the predominant view described bilingualism as a handicap, or at least as a mental burden that needed to be avoided (e.g. Weinreich 1953: 8), the current view ascribes to it much more positive properties. Most scholars consider bilingualism no burden at all, and even suggest that it may be beneficial to language development itself, as well as to other cognitive skills. For example, it has been claimed that bilingualism increases the ability to focus attention and that bilinguals are significantly better at filtering speech sounds from background noise (Krizman et al. 2012). Other research suggests that bilingualism may even delay dementia (Bialystok et al. 2012).

3.1.3.1 SECOND LANGUAGE ACQUISITION

Everyone who has taken on the challenge of learning a foreign language will have noticed that during the learning process elements from one language occasionally

certainty about results in this particular kind of setting. Although we may hypothesise that many principles apply in the same way to multilingual individuals, this needs to be tested explicitly.

slip into the other. In other words, keeping the two language systems separate does not always come naturally. The transfer may apply to lexical items as well as to grammatical structures, and may occur voluntarily or involuntarily depending on factors such as language proficiency and attitude of the speaker, as well as of the conversational partner. This leads to the question of how languages are organised in the brain. Does the bilingual brain contain separate systems for each language, or do all languages tap into one underlying linguistic and/or conceptual system?

Initially the assumption was that each language had its own self-contained structure and thus that the two linguistic and conceptual systems were entirely separated from each other (Penfield and Roberts 1959 cited in Hamers and Blanc 2000: 173). Current opinions reject such a radical separation, yet differ with respect to the extent to which the systems are assumed to be shared. While some scholars propose that the two languages share one system of conceptual representation (e.g. Meuter 2005: 349, Grosjean: 441), others suggest that the conceptual and linguistic systems are kept separate and are structured in a hierarchical way, with a common processor at the conceptual level and separate processors at the level of linguistic encoding (*ibid*). While the details of this issue remain unresolved, experimental research with bilinguals provides supporting evidence for the idea that there is interaction between the languages in the bilingual brain, and that even if they belong to different systems, they are certainly not isolated from each other.

For example, Meuter (2005) gives some insight into how this interaction comes about. She shows that once a language is being learned, it is always active and can never be 'switched off' (Meuter 2005: 352). Its activity level can be reduced, but a language is never deactivated entirely, which raises the question how a bilingual is able to choose a language without major interference from the other. Meuter demonstrates that this selection takes place through suppression of the undesired language rather than activation of the desired one³. This suppression is cognitively demanding, and increasingly so the greater the proficiency and familiarity with the language. When suppression is not completely

³ In her experiments with unbalanced bilinguals, Meuter shows that counter to expectations they take longer to switch from their L2 to their L1 than the other way round. Her proposed explanation for this pattern is that it is cognitively much more demanding to undo the strong inhibition that is needed to suppress the L1 than it is for L2, leading to the longer reaction time. This effect is only present in unbalanced bilinguals and does not show in individuals that are equally proficient in both languages.

successful, elements or structures of the undesired language may slip through the barrier and may be transferred into the other language. This explains how involuntary interference of the suppressed source language into the recipient target language may take place, and why interference from a bilinguals' dominant language (often L1), which is more difficult to suppress, into his non-dominant language (often L2) is more likely to happen than the other way round (see also Van Coetsem's theory in Section 3.1.3.3). Additional evidence for a porous boundary between the two language systems comes from work by Robinson and Ellis (2008: 7), among others, who show that the error pattern in L2 learners differs depending on the (dominant) L1 of the speaker, confirming the idea that languages in the bilingual brain are not isolated systems and clearly interact with each other.

While in situations of imperfect learning (e.g. due to insufficient exposure to the target language) the bilingual may resort to elements from his dominant language out of necessity, in other situations the bilingual may manipulate the degree of transfer to negotiate the right code with his conversation partner. This variation in relative activation level of the bilinguals' two languages is what Grosjean ([2000] 2007: 430) calls 'language modes'. Although in reality language activation is a continuum and cannot be divided into discrete categories, for descriptive purposes Grosjean identifies three modes, corresponding to three levels of relative language activation in a bilingual's languages A and B: a monolingual, intermediate and bilingual language mode. In monolingual language mode, a bilinguals' language A is highly activated, while for B the activation level is very low (i.e. it needs to be strongly suppressed). In the intermediate mode B is activated somewhat more, and in the bilingual mode languages A and B are both highly active, but A still serves as the 'base' language (Grosjean [2000] 2007: 430). Depending on the level of bilingualism of his conversational partner, the bilingual speaker will now decide which language to speak and how much interference from his other language is allowed, or even needed. This leads to different degrees of transfer in the speech of the same individual, conditioned by non-linguistic factors (i.e. conversation partner) rather than language proficiency. While Grosjean does not specify whether a certain language mode correlates with certain types of interference, Treffers-Daller (1998: 196) notes the following pattern of correspondence between language mode and linguistic reflection: in the monolingual mode there is only minimal interference in the form of some lexical copies; in intermediate mode interference is rather pragmatic in character, but the

switches between languages are clearly marked by pauses; in bilingual mode there are inter- as well as intrasentential switches between the languages, which occur without hesitation (Treffers-Daller 1998: 196-197). If it is correct to assume that these hesitations are the result of the effort it takes to conquer the inhibition of language activity, then this would mean that in the bilingual mode, where no delays were noticed in the speech flow, there is barely any suppression of either of the languages.

3.1.3.2 BILINGUAL ACQUISITION

While ‘mistakes’ made by second language learners are clearly a source of innovation, and thus a potential source of language change, alterations in the speech of bilingual first language learners can fulfil a similar function. In their longitudinal study of English-Chinese bilingual children, Matthews and Yip (2009) show that the two language systems of these children are also not isolated from each other during the acquisition process. Instead, the children match structures and semantic patterns in one language to patterns in the other through interlingual identification (see Section 3.1.3.3 for Weinreich’s description of this term in second language acquisition), leading to the innovative use of elements and constructions when the match is not perfect. Matthews and Yip focus in particular on the use of the English adverb *already* and the verb *give*, which the English-Cantonese bilingual children use as a perfective marker and as a permissive, respectively, based on the model of Cantonese. Matthews and Yip witness this in the synchronic speech of bilingual children, but importantly this use of these English elements is also a feature of Singaporean Colloquial English, a mixed language that developed as the result of contact between speakers of English and Cantonese. Of course, the bilingual children base their use of these elements on the synchronic polyfunctionality of corresponding Cantonese elements, and are by no means aware of the diachronic processes of grammaticalisation that underlie some of their uses in current Singaporean Colloquial English. However, the striking similarity between the innovative patterns of use in young English-Cantonese bilingual children and the established mixed language spoken by adults leads Matthews and Yip to the conclusion that bilingual acquisition may also be a significant impetus for the creation of contact-induced change. Most likely, the children in Matthews and Yip’s study will lose

their innovative use of English elements as soon as they go to an English-speaking school, but “...what would happen given a whole community of similar children?” (Matthews and Yip 2009: 390). In other words what would happen if there were nobody to ‘correct’ them? Or more specifically, what would happen if these children were not surrounded by speakers who follow the norms of Standard English, and who eventually serve as a model for them to revise their use of English? Matthews and Yip assume that this is exactly what happened in the development of Singaporean Colloquial English, thus showing how bilingual acquisition plays an important role in contact-induced change, alongside second language acquisition.

3.1.3.3 BILINGUALISM IN LANGUAGE CONTACT THEORY

As was mentioned in Section 3.1.1, the perception of the bilingual individual as the locus of language contact and his crucial role as a source of innovations in the onset of contact-induced change was recognised as early as Weinreich. While he clearly acknowledges the intricacy of each individual contact situation, the quotes below demonstrate that Weinreich assigned primary importance to the individual in his approach to language contact (1953: 71):

To predict typical forms of interference from the sociolinguistic description of a bilingual community and a structural description of its languages is the ultimate goal of interference studies. Unfortunately this aim cannot be attained till the missing link - the correlations between characteristics of individual bilinguals and interference in their speech - is supplied.

When one considers, however, that the bilingual speaker is the ultimate locus of language contact, it is clear that even socio-cultural factors regulate interference through the mediation of individual speakers.

Although Weinreich’s ideas reflect the predominance of structuralism at the time (see Section 3.1.1), which is why he is often mentioned in one breath with structuralists, my impression, based on quotes like the above, is that his view is much broader than this. While structural factors play a significant role in his ideas about contact-induced change, they seem to be only one of many factors that he considers necessary to a full understanding of the phenomenon. Thus, as a

linguist, one may “see the cause of the susceptibility of a language to foreign influence in its structural weaknesses” (Weinreich 1953: 4),

[b]ut the extent, direction, and nature of interference of one language with another can be explained even more thoroughly in terms of the speech behaviour of bilingual individuals, which in turn is conditioned by social relations in the community in which they live. In other words more complete findings can be expected from coordinated efforts of all the disciplines interested in the problems. (Weinreich 1953: 4-5)

Among these disciplines, he includes sociology, psychology, ethnography, pedagogy, geography and even law. The structuralist inheritance is apparent in Weinreich’s view that transfer may occur at the level of speech (or *parole*), corresponding to what is often called language variation, and at the level of language (or *langue*), corresponding to language change. Contact influence may manifest itself through the transfer of actual linguistic material, as well as through the important underlying process of ‘interlingual identification’ whereby the bilingual speaker matches linguistic elements or patterns of one language with elements or patterns of the other, even if that match is not perfect. In the latter case, interlingual identification may lead to changes in the phonology, syntactic structures or semantics of the speaker’s second language. According to Weinreich, this identification occurs in order to reduce the ‘burden’ of bilingualism since the assumption is that it is cognitively less costly to have a single set of linguistic distinctions to select from, rather than different sets for each language (but see Section 3.1.3).

A more recent theory that builds on the idea of the bilingual individual as the locus of contact is formulated by Van Coetsem (2000).

Second language acquisition is language contact. One can have language contact without second language acquisition, but not second language acquisition without language contact. Language contact, which occurs wherever one language is influencing another, is here a condition *sine qua non*. (Van Coetsem 2000: 35)

In his theoretical model, Van Coetsem distinguishes between two main processes of contact-induced change, identified as borrowing and imposition. The distinction between them is determined by the linguistic dominance of the bilingual speaker, whereby dominance is defined as the speaker’s relative proficiency in a certain language. Linguistic transfer is described in terms of a

‘source language’ (SL) and a ‘recipient language’ (RL), regardless of the material that is transferred (substance or structure) or the process of transfer (borrowing or imposition). Borrowing refers to a situation where the speaker transfers linguistic elements from his non-dominant SL (often L2) into his dominant RL, whereas imposition occurs when the speaker transfers linguistic features from his dominant SL (often L1)⁴ into his non-dominant RL. In this model, the dominant language is also always the ‘agentive’ language, either in projecting its own structures onto the non-dominant language, or by taking on (or receiving) features of the non-dominant language. Therefore, Van Coetsem associates imposition with SL-agentivity and borrowing with RL-agentivity. Schematically this looks as follows:

Table 3.1: Relation between linguistic dominance and processes of change according to Van Coetsem (1995, 2000)

Process	Linguistic Dominance		Agentivity	
Imposition	dom SL	→	non-dom RL	SL
Borrowing	non-dom SL	→	dom RL	RL

Another important concept in his theory is the stability gradient. According to Van Coetsem, certain language components are more stable than others, and are therefore less sensitive to influence from other languages. Importantly, the degree of stability in turn influences the kinds of linguistic transfer that may be expected in the processes of imposition and borrowing (Van Coetsem 1995: 67). In Van Coetsem's words, differences in stability reflect differences in ‘structuredness’ of linguistic domains as well as in ‘consciousness’ of the speaker about them (ibid.: 68). Stable parts of a language are linguistic domains that are more structured and consist of less independent constituents or elements. Since they are part of a system, the speaker is less conscious of the use of every individual element and their use is assumed to be more automatised during the speech production process. Van Coetsem's assumption is that this ‘automatic’ use of such language domains renders them less susceptible to change. The stable subsystems include syntax or the phonological and articulatory systems. Less stable parts, on the

⁴ In practice, the dominant language is often L1 and the non-dominant language L2, but this can change during a person's lifetime (Van Coetsem 1995: 70, 2000: 52). The concept of dominance has been discussed by various scholars and opinions differ as to whether dominance always correlates with L1 or whether a person's linguistic dominance can change over time.

other hand, include linguistic domains that have less internal structure and consist of more individual constituents. In contrast to the stable domains, a speaker's use of this part of his language is less automatised and more conscious, and therefore it is more prone to change. The most representative example of the less stable domain is the lexicon of content words⁵. In sum, the implication of the stability gradient for the transferability of linguistic properties between different language systems is that lexical items are more easily transferred than phonology and syntax (Van Coetsem 2000: 31).

Van Coetsem's idea that the dominant language is always the 'agentive' and active language during the process of linguistic transfer was empirically paralleled, and thus supported, by Meuter's finding that in the bilingual individual the dominant language is always hardest to suppress. The concept of the stability gradient suggests that in a situation where imperfect suppression obtains, the most stable (unconscious, automatised) components are most difficult to block, and are therefore most likely to percolate into, and eventually be transferred to, the non-dominant language. This could explain why during the process of imposition (where the SL is dominant) relatively stable syntactic and phonological properties are kept and transferred to the RL due to imperfect suppression of the SL, whereas during borrowing (where the RL is dominant) these stable domains of the RL remain relatively unaffected and lexicon changes first.

3.1.4. SOCIAL FACTORS

While knowledge of the psycholinguistic processes in second language acquisition and bilingual first language acquisition is indispensable for an understanding of linguistic innovations, it remains restricted to the explanation of variation on an individual level. For individual language variation to lead to change, innovations need to be disseminated and established in the (wider) speech community. In other words, the use of a particular variant needs to extend beyond the individual into the domain of the community, and the mechanisms at work for the establishment and propagation of this variant extend from psycholinguistic mechanisms in the human brain to social factors at the level of the community.

⁵ But see e.g. Bybee (1992: 70) and Brown and Witkowski (1983: 84) for a different opinion on structuredness of the lexicon.

These include degree of bilingualism, duration of contact, and the perceived prestige of the language variant. This section summarises the key ideas of some approaches that promote the idea that social factors determine the outcome of a contact situation, and may overrule the linguistic or psycholinguistic constraints discussed above.

3.1.4.1 INTENSITY OF CONTACT (THOMASON AND KAUFMAN)

The most prominent landmark in this domain is beyond doubt Thomason and Kaufman's 1988 book 'Language contact, creolization and genetic linguistics'. In contrast to the structuralists and generativists who had dominated the field for a number of decades, the main message of this work is that it is not the structural properties of the languages, or the innate properties of Universal Grammar, that determine the way in which languages can influence each other, but rather it is social factors that shape the linguistic outcomes of a contact situation. Thomason and Kaufman certainly do not deny the role of structural linguistic factors altogether. They acknowledge that they may facilitate or inhibit the establishment of a linguistic change, but they emphasise that social factors can always overrule these linguistic constraints:

It is the sociolinguistic history of the speakers, and not the structure of their language, that is the primary determinant of the linguistic outcome of language contact. Purely linguistic considerations are relevant but strictly secondary overall. Ultimately, all the proposed structural constraints discussed in chapter 2 fail because linguistic interference is conditioned in the first instance by social factors, not linguistic ones. (Thomason and Kaufman 1988: 35)

This implies that, counter to the line of thought pursued by the structuralists, everything can in theory be transferred under favourable social conditions.

As far as the strictly linguistic possibilities go, any linguistic feature can be transferred from any language to any other language; and implicational universals that depend solely on linguistic properties are similarly invalid. (Thomason and Kaufman 1988: 14)

Both the direction of interference and the extent of interference are socially determined; so, to a considerable degree, are the kinds of features transferred from one language to another. (Thomason and Kaufman 1988: 35)

This does not mean, however, that Thomason and Kaufman think that contact-induced change occurs in a random way. On the contrary, they were the first to propose an analytical framework of contact-induced change that aims at explaining, as well as *predicting*, the kinds of change to be expected in a particular contact situation (Thomason and Kaufmann 1988: 46). However, instead of linking the possible changes to purely linguistic constraints, they argue that certain linguistic outcomes are generated by certain social situations, and that from this the linguistic outcomes can be used to reconstruct past social situations.

Thomason and Kaufman recognise that the social factors in a contact situation are too many, and that their interaction is too complicated, to build a comprehensive predictive model of contact-induced change. Accepting this as a fact, they consider it worth establishing correlations between a single social factor and its linguistic outcomes. In the model proposed in their 1988 book, they focus in particular on intensity of contact, which appears to play a significant role in different kinds of contact situations that they discuss.

The main dichotomy Thomason and Kaufman make is between social situations of language maintenance versus language shift. The term 'maintenance' denotes a situation where "the native language is maintained but is changed by the addition of the incorporated features" (Thomason and Kaufman 1988: 37). The linguistic process associated with such situations is borrowing, defined by Thomason and Kaufman as "the incorporation of foreign features into a group's native language by the speakers of that language" (ibid: 37).

Shift, on the other hand, is a situation of language contact where speakers give up their native language, and adopt the language of the people they are in contact with (the so-called target language). In this situation it is changes brought about by the shifting speakers in the target language that are in the focus of interest. The linguistic result associated with this scenario Thomason and Kaufman call interference through shift, or substratum interference, which refers to the changes introduced by these second language learners through imperfect learning. This may be caused by insufficient access to the target language, attitudes of the shifting speakers, or because the shift took place very suddenly. As soon as this 'imperfect' version of the target language becomes established among

the original speakers of the target language, then substratum interference can be said to have occurred.

The difference between these two contact situations is not the type of change that is possible in the language. As was mentioned before, Thomason and Kaufman see no linguistic restrictions on what can be transferred. The difference is rather the order and the extent to which different types of change take place in a situation of maintenance or shift. In a situation of maintenance this depends on the intensity of contact, which they describe as a cline from light to moderate to heavy. In a situation of shift, this cline refers to the degree of interference in the recipient language, which in turn depends on the size of the shifting group and the level of bilingualism. Although Thomason and Kaufman's model is much more nuanced than I can possibly be in this summary, a sketch of their model will give an idea of the essential points. In a maintenance situation, casual contact and little bilingualism leads to copying of non-basic vocabulary into the recipient language. When contact is moderate to heavy, the number and range of copied lexical items increases and is supplemented by structural copies as well. In a shift situation, the likely sequence of change is reversed. In this situation, the phonology and syntax of the target language are affected first by the substrate language. The lexicon may be influenced to some extent too, but this influence is not as pervasive. Lexical items from the substrate language may enter the target language because this language has no labels for certain concepts, for example foods, tools etc., or if the target language is not easily accessible to the group of shifting speakers. Thomason and Kaufman (1988: 50) note that interference through shift only occurs when the shifting group is relatively large and does not learn the target language perfectly. If the group of shifting speakers is small and they blend with the target language speakers, or if they learn the target language perfectly, they do not leave any trace in the target language. Thus, while in a maintenance situation the lexicon of the recipient (native) language will change first (with potential influence in the grammatical structures later on), in a shift situation the phonology and syntax of the recipient (target) language will show the influence of the source language first (Thomason and Kaufman 1988: 50). While Thomason and Kaufman do not give a psycholinguistic interpretation of this observation, it is strikingly similar to what Van Coetsem explained using the notion of the stability gradient (see Section 3.1.3.3).

To recapitulate briefly, the idea behind the stability gradient is that phonology, morphology and syntax are more stable than the lexicon due to their

inherent structuredness and the less conscious use of these domains by speakers. Essentially, the more stable elements are given up less easily in contact situations. Applying this to Thomason and Kaufman's hypotheses, this could explain why in a maintenance situation, in which the dominant (often native) language is preserved, the stable systems of phonology, morphology and syntax typically remain free from large-scale restructuring, whereas the less stable domain of the lexicon is affected first⁶. In a shift situation the same principle of stability applies, but leads to the reverse effect. Like in a maintenance situation, the shifting speakers also preserve the most stable part of their dominant language (phonology, morphology and syntax), but unlike a maintenance situation, their adherence to these stable systems initially inflicts phonological, morphological and syntactic changes on the target language. The lexicon of the target language typically remains relatively unaffected. A schematic representation of the interaction between the social setting, psycholinguistic mechanisms and their linguistic outcomes is provided in Figure 3.1.

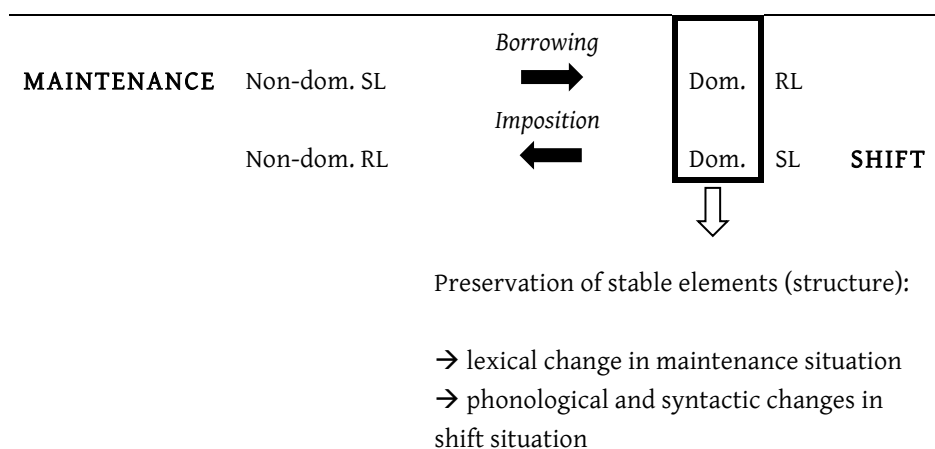


Figure 3.1: Interaction between social setting, psycholinguistic mechanisms and linguistic outcomes

⁶ But see Section 3.1.4.2 for a discussion of metatypy.

3.1.4.2 EMBLEMATICITY (ROSS)

While Thomason and Kaufman's model covers a large number of contact-induced changes, it does leave an important phenomenon unaccounted for. This is the situation where large-scale morphosyntactic restructuring takes place in a setting of language maintenance in a bilingual community. This process was labelled 'metatypy' by Ross and defined as

[a] diachronic process whereby the morphosyntactic constructions of one of the languages of a bilingual speech community are restructured on the model of the constructions of the speakers' other language. (Ross 2007: 116)

This new term was necessary to describe a contact setting not accounted for by Thomason and Kaufman's model, namely the structural reorganisation of a language in a maintenance situation. While in their model some space is reserved for structural borrowing, this does not represent the fundamental and system-wide typological change that may occur in contexts of long-term bilingualism (Ross 1996, 2007).

In metatypy, there is very little transfer of actual linguistic material between the languages in contact. Instead, the morphosyntactic or semantic organisation of the source language is transferred to the recipient language on a system-wide scale. As a result, the morphological material of the recipient language may look like its genealogical relatives, but its morphosyntactic and semantic structure, or 'type', may be closer to the contact language. This system-wide restructuring sets metatypy apart from phenomena such as grammatical calquing and borrowing, although these processes appear to precede metatypy and the boundary between them cannot always be clearly drawn (Ross 2007: 132).

The process can be illustrated most clearly with Ross's own, now classic, example of Takia, an Oceanic (Austronesian) language spoken on Karkar Island in Papua New Guinea, which has been restructured on the model of Waskia, an unrelated Papuan language spoken in the same area. The result is a word-by-word structural correspondence between Takia and Waskia, while the language material has clearly remained different. Example (3.1) shows that Takia differs in structure from Arop-Lokep, a closely related Oceanic language, and that it has adopted exactly the same structure as the genealogically unrelated language Waskia:

- (3.1) Arop-Lokep: *tool tamoto ma rima-na*
 man male and wife-his
 ‘a man and his wife’
- Takia: *ŋai tamol an ida*
 Waskia: *ane kadi mi ili*
 1.SG man DET with.him
 ‘The man and I’ (Ross 2007: 121)

It is obvious that this kind of large-scale restructuring can only happen in a situation of intense contact. However, the curious phenomenon that such intense contact has not led to concomitant lexical copies from Waskia into Takia, or to a situation where Takia speakers shift to Waskia, is often explained by assuming a high degree of ‘emblematicity’ of the metatypised language (Takia) (Thurston 1987, Crowley 2000, Ross 2007). In such cases, language serves not only the purpose of communication but also functions as an ‘emblem’ of a speech community. In other words, the native language carries high identificational value and is an important marker of identity. Therefore the native language is not given up easily in such settings and despite intense contact large-scale borrowing or shift is unlikely to occur. Metatypy, on the other hand, causes fewer problems, since the language forms (and thus the audible, and identificational part) are largely maintained, and only the structure is changed on the model of the contact language. Thus, the social situation associated with metatypy is a bilingual speech community, in which “speakers of the emblematic language(s) are a significant majority and where a variety of the inter-community language is used at least as much for communication among those speakers [i.e. speakers of the emblematic language E.S] as it is for communication with speakers of another language or of other languages” (Ross 2007: 131). According to Ross (1996: 182, 2003: 183), it is typically the emblematic language of a bilingual speech community that undergoes metatypy, but the other way round is also possible (2007: 130). It has been suggested that this process is motivated by the urge of bilingual speakers to use only one cognitive representation of a certain proposition (Sasse 1985: 84-85, 1990: 32, Ross 1996: 204, 2007: 134-135), and reduce the effort it takes to match meaning to form (see also interlingual identification 3.1.3.3).

3.1.4.3 SOCIO-SPATIAL FUNCTION AND ATTITUDE (THURSTON, ANDERSEN, WRAY & GRACE)

Thomason and Kaufman's model offers clear and straightforward correlations between intensity of contact and its linguistic outcomes, but they admit that their analysis of other social factors such as the degree of bilingualism, perceived prestige of the language variants, and attitudes of the speakers remains superficial. They justify this by saying that "these are as varied as the contact situations in which they are embedded" (Thomason and Kaufman 1988: 46), and therefore they consider it unrealistic to implement them into their model in any schematic way.

While it is true that social factors in a community are extremely complex, diverse and often unpredictable, other scholars have ventured into this domain, and tried to find regularities in the social mechanisms and group dynamics that underlie internally, as well as externally, motivated language change. Their research focuses in particular on the function of different languages within a speech community, and on how speakers consciously or unconsciously manipulate their language(s) to facilitate or impede communication. Thurston formulates this as:

People create new languages to be able to communicate with outsiders and change their languages to create barriers between themselves and their neighbours.
(Thurston 1989: 577)

With such statements, Thurston refutes the concept of language as an autonomously evolving organism that diversifies internally up to the point that a new language emerges. To model this idea, Thurston distinguishes between exoteric and esoteric languages. Exoteric languages are used for wider communication. They are meant to include as many speakers as possible in the group of speakers, the typical example being a lingua franca. They are used for inter-group communication and need to be easily learnable by in-group as well as out-group members. Therefore exoteric languages tend to be, or to become, structurally more simple and regular. The reverse applies to esoteric languages. Their primary sociolinguistic function is in-group communication, and according to Thurston, often even to impede communication with out-group members, thus making it an emblem of the speech community. In contrast to exoteric languages, the phonological, morphological and lexical systems of such languages tend to

become more complex over time, and consequently more complex for outsiders to learn.

A similar idea is expressed by Andersen (1988: 74) when he talks about open and closed dialects⁷. While Thurston's notions of exoteric and esoteric incorporate both the socio-spatial function of a language variety and the attitudes of the speech community towards language norms, Andersen prefers to separate these two aspects, using different sets of terms; namely 'open' versus 'closed', and 'exocentric' versus 'endocentric'. The opposition 'open' and 'closed' refers to the difference in socio-spatial function of languages. 'Open' languages are spoken in a community that maintains many relationships with other communities, whereas a 'closed' language is spoken in a community that has limited external links. The distinction between 'exocentric' versus 'endocentric' reflects differences in speaker attitude. In exocentric communities (also 'loose-knit' or 'outward-facing' elsewhere in the literature; see Ross 2003: 179, Wray and Grace 2005: 549, respectively), speakers have an accepting attitude towards influence from foreigners, including their languages, whereas 'endocentric' communities (also 'tight-knit' or 'inward-facing'; Ross 2003: 179, Wray and Grace 2005: 549, respectively) are less open to change of their existing language traditions. Thus, in Andersen's opinion, it is often not the actual amount of interlingual communication, but rather the collective attitude of a community towards language norms that has the most profound influence on the development of a language variety (Andersen 1988: 74).

Departing from the distinctions formulated by Andersen and in particular by Thurston, Wray and Grace (2005) further elaborate the notions of exoterogeny and esoterogeny. Like Andersen, they connect them to a particular social settings and speaker attitudes, but, innovatively, they also associate them with specific linguistic consequences. Exoteric languages are compositional, have transparent rules, and are regular and simple. In contrast, esoteric languages tend to be more formulaic, irregular and morphosyntactically opaque, which means that they must be stored in the brain in larger units, and are therefore harder to learn. According to Wray and Grace, these different ways in which languages develop is due to differences in socio-cultural dynamics. Esoteric languages can 'afford' to be formulaic, less transparent and less compositional because they are often spoken in communities in which the speakers have a high degree of shared knowledge.

⁷ Andersen's classification is based on the study of dialects, rather than languages, but he points out that the same principles apply for (un-)related languages (Andersen 1988: 75-76).

Wray and Grace even go as far as calling many of the actions and events in such communities ‘predictable’, in which case it is unnecessary to be overly explicit about everything linguistically. By contrast, the large number of out-group ties in exoteric communities requires detailed and explicit linguistic encoding, compensating for the lower degree of shared knowledge among speakers. Another important point that Wray and Grace draw on is the differing learnability requirements in these two kinds of communities. The inward-facing attitude of esoteric communities and their low number of links with outsiders imply that their language is unlikely to be learnt by many adult second language learners. This means that it need not fulfill the requirements of learnability demanded by highly analytic adult minds, and so not necessarily develop the analytic and regular structures that are likely to be introduced by adult second language learners. The language of an esoteric community is learnt mainly by children and as a first language. Since children are able to store language chunks in much larger units than adults, it is not a problem for more formulaic and irregular language structures to be maintained in these languages (Bolinger 1975: 100, Diessel and Tomasello 2001: 134, 136, Wray 2002: 105). The opposite is true for exoteric communities. The greater number of out-group links in this type of community implies that adult second language learners are a common phenomenon, and that native speakers have to adjust their ways of speaking in order to communicate with non-native speakers. This may lead to the development of more simple and regular structures. Equally, the language may be simplified and regularised by the second language learners themselves, as for example in a *lingua franca*.

3.1.5. FROM INNOVATION TO CHANGE

The previous two sections have shown that both the psycholinguistic mechanisms in the bilingual individual and the social setting in which a new language variant is embedded are essential factors for the understanding of contact-induced change. Considerable progress has been made in the study of these realms separately, but the connection between them has remained relatively unexplored. In other words, how does an innovation in the speech of an individual come to be established as a change within a speech community? This question applies to internally motivated as well as to contact-induced change, and is what Weinreich describes as the actuation problem:

Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times? (Weinreich et al., 1968: 102)

The answer to this question can clearly not be reduced to purely structural or psycholinguistic factors. Despite the lack of a clear-cut solution, promising insights into this issue are provided by social network theories, which became known most notably through the work of Milroy and Milroy (1985). In this article and similar studies, network theory is applied to account for the emergence and spread of language-internal innovations (Labov 1980, Milroy and Milroy 1985), but there is no stringent reason to assume that different principles apply in contact-induced change.

Another important factor that has been proposed to influence the origination and propagation of change is speaker age (Kerswill 1996, Ross forthcoming). It appears that certain types of change are associated with particular age groups and are unlikely to occur in others. Therefore age of speakers plays a significant role in the understanding of contact-induced innovations and change within a speech community.

3.1.5.1 NETWORK TIES (MILROY AND MILROY)

According to Milroy and Milroy (1985: 364), the emergence and propagation of linguistic change is governed primarily by the number and strength of the social network ties of an individual. They propose that weak social ties are crucial in the initial stages of an innovation, whereas strong network ties play a role in the propagation and establishment of the innovation, which could eventually convert the initial variation into an established change in the speech community. Milroy and Milroy elucidate this as follows. Weak ties are associated with superficial connections between individuals and they typically serve as ‘bridges’ *between* different social groups. A corollary of this superficiality is that a single individual can maintain a large number of such ties, but they have low identificational value. That is, weak ties are typically associated with individuals that occupy a marginal position in the social network, do not strongly conform to the social norms, and are therefore more open to innovation. The opposite is true for strong ties, which typically occur *within* social groups. Their intense character limits the number of

strong ties that a single individual can maintain, but such connections carry high identificational value. They are associated with individuals who occupy a central role in the network and that are typically norm reinforcing.

Transposing this to the domain of linguistics, this means that individuals with strong ties tend to use their language variety as a marker of the network to which they belong, and thus of their social identity. Since language is thus tied to identity and group membership, the conservative and norm-reinforcing character of strong ties makes speakers with such connections less likely to introduce linguistic innovations. It will be clear that weak ties are characteristic of inter-group relations, whereas strong ties promote cohesion within a social group. In contrast, the open character of weak ties and their association with inter-group relations, as well as the sheer number of them that an individual can maintain, establish the right conditions for the spread of language-internal, as well as for contact-induced innovations.

Against this background, Milroy and Milroy propose that initial innovations often occur in marginal individuals with many weak ties, who they call the innovators (Milroy and Milroy 1985: 367). However, for an innovation to spread through the community it is dependent on central figures with a network of strong ties, who they call the early adopters. Since these individuals are often personalities with strong identificational value for the group they are able to establish a new norm. Such speakers' use of a new language variant is crucial for its propagation and thus its establishment as an accepted variant across the speech community⁸.

Purely linguistic factors, the Milroys argue, may limit the possible range of innovations, yet they are not sufficient to explain patterns of diffusion of an innovation, and so to solve Weinreich's actuation problem. Occasionally linguistic factors may explain why a certain change took place (e.g. to avoid homophony) but they give no clue as to why other changes have not occurred, (e.g. instances of homophony that were not avoided; Milroy and Milroy 1985: 345). The answer to such questions must be searched for in psychological and social factors instead.

⁸ This differs from Labov's idea that the innovators are always people considered to have high prestige and with strong ties both inside and outside certain networks (Labov 1982).

3.1.5.2 AGE

Kerswill concludes with respect to this issue that everybody can, and does, continuously modify their language throughout their life, but *what* individuals are able to change and adopt in their language variety is to some extent determined by age (Kerswill 1996: 178). Based on the acquisition of English and Norwegian dialects, Kerswill proposes a hierarchy of dialect features relating to their difficulty to acquire. In this hierarchy, unpredictable phonological rules, new phonological oppositions and certain grammatical changes are the most difficult to learn and are only acquired early in life, varying from the age of 3 for unpredictable phonological rules to 13 for new phonological oppositions.

This is followed on the hierarchy by prosodic systems, which must be acquired before the age of 12-15 for a perfectly native-like performance, and by new morphological classes, the acquisition of which peaks during adolescence. Other changes such as morphologically conditioned changes, reanalysis, mergers, Neo-grammarians sound changes, lexical diffusion of phonological changes and the introduction of new vocabulary can be acquired during an individual's entire lifespan. Thus, children seem to be able to make any kind of change to their speech, whereas adults are much more restricted.

However, while children have the ability to take on any kind of innovation, in practice this does not happen on a large scale. Small children tend to imitate their parents closely, and the first divergence from this pattern appears in preadolescence (6-12 years), when the child's way of speaking comes to be more influenced by his or her peers than by the parents as he/she develops network ties beyond the core family. Despite the large potential for innovation in preadolescents, Kerswill concludes that it is adolescents who play the greatest role in the diffusion and establishment of new language variants.

Adolescents are clearly significant bearers of change; their networks allow them to have wider contacts than younger children, and their desire for a distinct social identity means that they are willing to modify their speech. (Kerswill 1996: 198)

A similar, but slightly different conclusion, is reached by Ross (forthcoming) in his discussion of the role of speaker age in grammatical restructuring. In accordance with Kerswill, Ross proposes that this kind of change is propagated primarily by adolescents. However, more explicitly than Kerswill he identifies preadolescents as the source of innovations (Ross forthcoming). Thus, while preadolescents can be

equated with Milroy and Milroy's innovators, the role of the early adopters is fulfilled by adolescents. This association of changes with certain age groups can be useful for the interpretation of contact-induced change and conversely for the reconstruction of the social situation in which it emerged.

3.1.6. SUMMARY

The previous sections have shown that factors from at least four different domains need to be considered for an explanation of contact-induced change. These domains, the factors and the hypothesised linguistic outcomes that were discussed in the previous sections are summarised in Table 3.2.

Table 3.2: Factors influencing linguistic outcomes of a contact situation

Domain	Factor	Hypothesised linguistic outcome
Linguistic structure	-Structural	Independent elements are transferred more easily than bound morphology
	-Typological	Grammatical structures can be transferred only between typologically similar languages
Bilingual individual	-Stability	Structured, automatised language parts (phonological, morphological, syntactic structure) tend to be more stable than the lexicon, and are therefore retained longer in contact situations.
	-Linguistic dominance	Dominant language is hard to suppress: RL dominance → borrowing (lexicon first) SL dominance → imposition (phonology and syntax first)
Bilingual society	-Intensity of contact	Influences the kinds and order of contact-induced change that tend to occur in situation of maintenance or shift
	-Emblematicity	May lead to metatypy
	-Function and attitude	Closed, tightknit communities → complexification Open, loose-knit communities

		→simplification
Position in society	-Network ties	Weak network ties → innovation Strong network ties → propagation of change
	-Age	Preadolescents and adolescents play a crucial role in innovation and change in contact situations

- 1) The structural and typological properties of the languages in contact: within the structuralist tradition the dominant view was that the structures of the languages in contact determine what elements can be transferred across the languages. While it may be true that transfer occurs more easily of morphologically independent elements and between typologically similar languages the number of counterexamples suggest that other factors can overrun these structural constraints.
- 2) The psycholinguistic processes in bilingual speakers with respect to both L2 acquisition and bilingual L1 acquisition were shown to play an important role in kinds of contact influence that are attested in their speech. In particular, language dominance appears to be crucial in the distinction between the processes of borrowing and imposition, and their linguistic consequences (Van Coetsem). It was discussed how language selection in bilinguals is dependent on the relative activation level of languages, and it was proposed that transfer takes place in cases of insufficient suppression of the dominant language (Meuter). This may happen involuntarily, depending on proficiency and/or the emotional and mental state of the speaker, or may be managed voluntarily depending on the partner in conversation (see Grosjean's language modes). In case of insufficient suppression of the dominant language, the stable (i.e. phonological and morphosyntactic) components of a language are preserved the longest. This leads to primarily lexical transfer during the process of borrowing, where the phonological and morphosyntactic systems of the dominant recipient language are preserved, and to phonological and syntactic transfer during imposition, where these systems of the dominant source language are copied into the recipient language.
- 3) Social factors are of great importance and can overrun the structural factors proposed under 1). In particular, intensity of contact was discussed, which influences the kinds of contact induced-changes, which tend to occur in a

situation of language maintenance or shift. In addition, emblematicity was identified as a factor that can prevent a community from shifting their language despite intense contact. This may lead to large-scale restructuring (or metatypic changes) in the emblematic language, yet with retention of the native lexicon. Finally, the socio-spatial functions of languages in the community (open vs. closed) as well as the attitude towards community norms (loose-knit vs. tight-knit) were discussed. It was claimed that language change in closed and tight-knit communities would lead to complexification and formulaic language use which is hard to learn for outsiders, whereas changes in open and loose-knit communities tend to simplify the system due to frequent interaction with outsiders.

- 4) Position in society (network ties) and speaker age were proposed to give insights into the processes by which innovations spread. Innovators were characterised as individuals with many weak network ties, while the propagation of an innovation is carried out by early adopters, who are central figures in the community with stronger, but fewer network ties. The acquisition of different kinds of linguistic variants was shown to depend on speaker age (Kerswill 1996), and it was suggested that for some kinds of change, at least, preadolescents are the likely innovators, while the role of early adopter is fulfilled by adolescents.

3.2. THE ROLE OF THEORY IN LANGUAGE CONTACT STUDIES

The theoretical approaches summarised above provide only a snippet of the available literature, yet their diversity shows that scholars in the field of contact linguistics are still far removed from a consensus on one all-encompassing theory of contact-induced change. This is as confusing as it is understandable, given the complex character of contact situations, and it is questionable whether it will ever be possible at all to formulate a single all-encompassing model which explains and predicts contact-induced change (Thomason 2010: 33). Since this is an unsatisfactory situation in research, this section will illuminate some of the reasons and justify to some extent why the study of contact-induced change is incompatible with the requirements of a predictive scientific theory in the strictest sense.

Most generally, a theory is intended as a model of reality. However, their hypothetical nature implies that theories can always be falsified and never be proven. Therefore we can never guarantee with certainty that a theory correctly reflects reality. Nonetheless, some theories are more robust than others. It has been proposed that a 'good' theory should be: a) explanatory; b) predictive; and c) testable (Hawking 1988: 11). The formulation of such a theory is always a demanding task, but the multifaceted character and non-reproducible nature of social and linguistic contact situations make the formulation of an all-encompassing theory particularly challenging within this domain.

It will be clear by now that the discipline of 'language contact' is in fact itself multidisciplinary. Since language is an important means for linking a person's inner world to the outside world, it cannot be emphasised enough that the underlying motivations for the linguistic outcomes of a contact situation cannot be fully understood unless psychological, sociological, anthropological, and historical aspects are taken into account besides linguistic factors.

Both internal and external motivations are needed in any full account of language history and, by implication, of synchronic variation. Progress in contact linguistics depends, in my opinion, on recognizing the complexity of change processes — on resisting the urge to offer a single simple explanation for all types of structural change. (Thomason 2010: 31)

The paradox of the desire for a theory of contact-induced change is that the scientifically uncontrollable interplay of all these factors is in disagreement with the nature of theories themselves to represent reality in a schematic and predictable way. This problem applies in particular to the aforementioned criteria of predictability and testability. To achieve these criteria, a clear cause and effect must be identified in the observation for which the theory accounts. However, as we have just seen in the quote from Thomason, the linguistic outcomes of a contact situation often require an explanation in terms of multiple causations and the sheer number of potential factors makes prediction of the way they will interact and the exact repercussions this will have on language an unrealistic goal (Thomason 2001: 61, 2003: 709, 2010: 33).

Another issue is the variable nature of the operating factors. While some tend to be stable (e.g. morphological paradigms), others have the potential to change within a brief period of time. This applies to important factors such as speakers' attitude towards a language variety, which may be affected by circumstances such

as social, economic or political development, or even by the appearance of a single prominent individual, and is therefore inherently unpredictable.

Conditions with respect to testability are not any more favourable. In order to corroborate the validity of a theory, it is necessary to test hypotheses through replication of the study. Ideally, one should be able to control the setting of the study and vary individual factors to see which factor is decisive in the emergence of the observation under consideration. However, in language contact the attainment of both these goals is unrealistic. First, no two naturally-occurring contact situations are exactly the same, and it would be unethical, as well as technically impossible, to create them artificially. Therefore, it will never be possible to perform an exact replication of a study to test if the linguistic outcomes would be same. But even if it were possible to do this, the problem remains that one can never be entirely sure that all relevant details of the contact situation were recorded, or even known. Even if a theory makes sense on the basis of the available information, we cannot be sure that some other crucial factor that we have missed did not also play a role in the linguistic outcomes. Finally, we should not forget chance. Some linguistic changes to which we assign theoretical significance may in reality have occurred by chance (Butters 2001: 201).

Given this background, the conclusion among contact linguists that it is unrealistic to aspire to the criterion of predictability is no surprise. Rather, one should accept the fact that contact-induced change can only be observed and explained after the event. This is well formulated by Thomason:

In spite of dramatic progress toward explaining linguistic changes made in recent decades by historical linguists, variationists, and experimental linguists, it remains true that we have no adequate explanation for the vast majority of all linguistic changes that have been discovered. Worse, it may reasonably be said that we have no full explanation for any linguistic change, or for the emergence and spread of any linguistic variant. The reason is that, although it is often easy to find a motivation for an innovation, the combinations of social and linguistic factors that favour the success of one innovation and the failure of another are so complex that we can never (in my opinion) hope to achieve deterministic predictions in this area. [...] The realistic goal is a deeper understanding of processes of change, not an ultimate means of predicting change. (Thomason 2010: 33)

However this does not mean that the field of contact linguistics cannot be advanced at all. While the desire to formulate strict *laws* of contact-induced change may be too far fetched, there is certainly agreement on *tendencies* as to

what is more or less likely to happen in particular contact situations. That is, a theory of contact-induced change should be able to explain, but not necessarily be able to predict, the outcomes of a contact situation.

To conclude, theories in language contact studies have a status of their own and should not be compared to theories in the strictest sense of the term, which requires the fulfilment of the three criteria mentioned above. Explanations and predictions should be formulated in terms of ‘probabilities’, rather than ‘laws’. Theories of language contact should be seen as guidelines in the reconstruction of social settings in the past, as well as in the prognosis of probable outcomes of contemporary contact situations, but not as a rigid framework that does not allow for exceptions.

3.3 TERMINOLOGY

The diversity of theoretical approaches in language contact studies is mirrored by an equal diversity of terminology. Therefore I will specify the set of terms and their meanings that have been chosen for use in this dissertation.

The term *COPY* is used to refer to any kind of linguistic material that is inserted in one language from another (see Johanson 1992: 175, 2002a: 8, 2002b: 287). Within this category, *FULL COPIES* denote form-meaning units that are adopted wholesale into another language (e.g. concrete morphemes and their meanings), whereas *COPIES OF FORM* and *COPIES OF STRUCTURE* represent cases where only a subset of the properties of the item in the source language is copied into the recipient language. Copies of form refer to the insertion of linguistic forms (or substance) from the source language into the recipient language, whereas copies of structure designate cases where only structural features (and no linguistic forms) are copied. Structural features may refer to semantic, morphological or syntactic structure, as well as frequential or combinatorial properties of linguistic items⁹.

Regarding the processes involved in contact-induced change, the most neutral labels to refer to the spread of linguistic material from one language to another is *COPY* or *TRANSFER*. Regardless of the nature of the copied material or of

⁹ Cf. Johanson’s (1992, 2002) classification of ‘global’ vs. ‘selective’ copies, whereby global copies largely overlap with what I call full copies, and selective copies include both copies of form and copies of structure.

the underlying process, transfer always takes place from a SOURCE LANGUAGE into a RECIPIENT LANGUAGE (Van Coetsem 2000: 49).

However, it is more informative to categorise copies based on the prevailing sociolinguistic and psycholinguistic factors in the contact setting. Following Van Coetsem, I will use BORROWING for the process where linguistic material is copied into the recipient language due to recipient language agentivity, and IMPOSITION when this happens due to source language agentivity. The difference in agentivity is determined by a difference in linguistic dominance of the bilingual speaker: in borrowing, the speaker's dominant language is the recipient language, and copies are transferred into this language from their non-dominant source language. In imposition, on the other hand, the speaker's dominant language is the source language, and copies are transferred into the speaker's non-dominant recipient language.

Linguistic DOMINANCE is defined as the difference in the bilingual speaker's degree of proficiency between his two languages (Van Coetsem 2000: 32). While some authors equate linguistic dominance with L1 automatically (e.g. Lucas 2012: 277-278), I agree with Van Coetsem's idea that in certain situations a person's relative proficiency, and thus his dominant language, may change during his lifetime (for example, migrants who have left their own country and subsequently use the language of their new homeland (their L2) much more than their mother tongue (L1). If such a situation continues for long enough, the L2 will become much more active in the brain than the L1, and the speaker can legitimately be called dominant in his L2.) In addition to borrowing and imposition, there is METATYPY, which refers to a situation where a language, often emblematic for the speech community, undergoes large-scale restructuring as a result of long-term bilingualism and intense contact with another language.

One final relevant process that deserves mention is language ATTRITION. While there is some difference in its definition, there is general agreement that

attrition is a gradual process in which a language recedes as it loses speakers, domains, and ultimately structure; it is the loss of linguistic material that is not replaced by new material. (Thomason 2001: 227)

Attrition is also a part of contact-induced change, but its character is different from borrowing and imposition in that the linguistic consequences are indirect. The changes do not necessarily make the affected language more similar to the other language and there is not necessarily transfer of substance or schematic

copies. However, the changes would be less likely to occur outside the contact situation (Thomason 2003: 688).

Attrition occurs when a language loses ground as the result of a shift to a different language by a speech community, and it may eventually lead to language death. The loss mentioned above applies to all linguistic domains and the changes are characterised by simplification of the system of the dying language. Loss of vocabulary is observed in certain domains, often as a corollary of the loss of cultural practices or traditional knowledge. Therefore it tends to start in marginal domains and gradually proceeds to the core lexicon of the dying language. The most characteristic symptom of attrition, however, is structural simplification. This applies to the phonological system as well as to morphology, syntax and even discourse structure. This often comes to the fore as generalisations of rules, leading to a reduction of irregularities, the merger or elimination of phonological or morphosyntactic categories, the replacement of morphologically complex by analytical constructions, and the loss of complex syntactic constructions.

It is not easy to prove whether a change is the result of attrition or whether it is simply a language-internal development. Lexical loss and replacement by foreign lexical items also occur in situations of borrowing, and structural simplification is also a natural tendency in languages that show no signs of attrition.

Therefore before any generalisations are made, every situation needs to be treated separately, carefully taking into account the social situation, and leaving open the possibility of multiple causation in the explanation of contact-induced change (Thomason 2001: 232).

4.1 INTRODUCTION

One linguistic domain in which contact influence can easily be recognised is the lexicon. Lexical items of foreign origin often stand out in their phonological or morphological structure and are therefore readily identifiable as copies from another language. However, this only applies to so-called full copies (see Section 3.3), where the complete unit of form and meaning are copied into the recipient language. Apart from these cases, there are many examples where either only the phonological form, or the meaning of a foreign lexical item is transferred to the recipient language. In the latter case this means that the phonological form of a word remains the same, while its semantic structure become more similar to the semantic structure of the source language. This phenomenon has been referred to by a variety of terms in the literature, including *Lehnbedeutung* ‘semantic loan’ (Blank 1997: 349, Betz 1949: 15), loanshift (Smith-Stark 1994: 17), lexical calquing (Ross 2007: 122)¹ or semantic borrowing (Geeraerts 2010: 29). Since it is impossible to identify this type of semantic transfer on the basis of the lexical form in the recipient language, the only way to detect changes of this kind is to make a fine-

¹ In earlier work Ross referred to this phenomenon as lexical metatypy (Ross 1996).

grained comparison of the semantic organisation of relevant semantic fields in the source language (SL) and the recipient language (RL).

In this chapter, lexical differences between Dolgan and Sakha are investigated, including both full copies and structural copies, with particular emphasis on instances of semantic change. The attested differences are analysed from a quantitative (4.4.1) as well as from a qualitative point of view (4.4.2) and in cases where a semantic difference between Dolgan and Sakha is attested it will be investigated whether its emergence could have been motivated by contact.

Since the lexicon is only one of the linguistic domains that need to be examined with respect to potential contact-induced change, the aim of this chapter is not to give an exhaustive inventory of differences across the entire lexicon. Rather, it focuses on a limited but comparable set of meanings in Dolgan and Sakha, allowing for an in-depth analysis of the quantitative as well as qualitative properties of the attested differences.

4.2 LEXICAL CHANGE AND LANGUAGE CONTACT

Most of the literature on contact-induced change in the lexicon focuses on the transfer of lexical copies from a source language into a recipient language. By default the tacit implication is that full copies are concerned, i.e. both phonological form as well as semantic structure are imported into the recipient language as a single unit. However, data from language contact studies show that this assumption is not always correct and it appears that the two aspects of De Saussure's linguistic sign can be, and often are, transferred separately. As mentioned above, the signified, corresponding to the semantics of a lexical item, may be copied independently from the signifier (or phonological form) into the recipient language, and can be projected onto existing native phonological forms, leading to a semantic change in this language. Although there is a comparatively small body of literature on this subdomain of lexical change and the sociolinguistic conditions underlying the phenomenon, it has been recognised and described by different scholars, and before presenting the data from Dolgan, a brief overview of relevant work in this domain is in order.

The categorisation of lexical change in most current theories is founded on work by Betz (1949) and Haugen (1950, 1953). Betz made a distinction between *Lehnwort* 'loanword' and *Lehnprägung* 'loan coinage' or 'loan meaning', whereby

Lehnwort refers to the transfer of both form and meaning (i.e. a full copy), and *Lehnprägung* to the transfer of meaning only (i.e. semantic copy). An example of the first category is the word *curry* in English. Here both the concept of ‘spicy sauce’ and the phonological form *curry* were adopted from Tamil *khari* ‘sauce’. An example of the second category is the extension of the word *frio* ‘cold temperature’ in American Portuguese to include the additional meaning of ‘cold infection’ (Winford 2010: 172). This category also includes calques (or loan translations) of the type ‘skyscraper’, compounds of which the semantic structure has been copied into several languages, e.g. German *Wolkenkratzer*, Dutch *wolkenkrabber*, Russian *neboskreb*, French *gratte ciel*. This distinction was further refined by Haugen (1949: 289-290, 1950: 219-220), who subdivided Betz’s category of *Lehnprägung* into loanshifts, including loan meanings (i.e. semantic copies) as illustrated above, and creations. A creation is defined as an innovative combination of native lexical items that is not directly copied from the source language, but is somehow inspired by the contact situation. An example is the Pima expression ‘wrinkled buttocks’ to refer to the concept of elephant that was foreign to the Pima speakers (Winford 2010: 173). In this case, the concept was introduced through contact, but the lexical form or semantic structure is not. Thus Haugen’s loanword (i.e. full lexical copy) is defined as the copy of the unit of phonological form and semantic structure into a recipient language, whereas loan meanings (i.e. semantic copies) are “changes in the semantics of an RL word under influence from an SL word”, (Winford 2010: 172). As mentioned above, later scholars have treated the latter phenomenon under the names of loanshift (Smith-Stark 1994: 17) or semantic borrowing (Geeraerts 2010: 29). The crucial difference between full lexical copies on the one hand and semantic copies on the other is that only for full lexical copies is there transfer of actual phonological form. For semantic copies (including purely semantic copies as well as calques) semantic structures are transferred from a source language into a recipient language, but these are mapped onto a native phonological form.

Most theoretical models of language contact differentiate between the copying of phonological form and the copying of grammatical or semantic structure (see Chapter 3). They make distinctions such as global vs. partial copying (Johanson 1992: 175), copying of matter vs. pattern (Matras and Sakel 2007: 829-830, 2008: 15), substance vs. schematic copies (Croft 2003a) or diffusion of form vs. diffusion of pattern (Aikhenvald 2003: 3). However, while some scholars would associate these different kinds of linguistic change with different socio-historical

and sociolinguistic conditions (e.g. Thomason and Kaufman 1988), different kinds of change within the lexical domain are rarely linked to different social conditions. One model that accounts for this distinction is the work by Ross in his discussion of metatypy, a term he uses to describe the restructuring of the recipient language on the model of a source language, whereby “the ‘metatypised’ (restructured) language maintains forms resembling those in its genetic relatives but the meanings of these forms have changed” (Ross 1996: 182). This restructuring may affect the grammar or morphology of the recipient language, but may equally well apply to the lexicon, in which case the phenomenon is referred to as lexical metatypy (Ross 1996: 191). To put it even stronger, Ross (2007: 124) seems to imply that morphosyntactic metatypy originates in lexical metatypy, which thus could be the source of more large-scaled restructuring of a language. He argues that semantic reorganisation develops out of different construals of reality that individuals are faced with while learning different languages, imposing an extra cognitive burden on the brain. Bilingual individuals “shed much of the burden by bringing the semantic organizations and reality construals of their two languages into line with each other” (Ross 1996: 204). They do that first by assimilating the ways of saying things in one language to the semantic organisation of their other language, which may then be extended to the restructuring of morphosyntactic categories.

Although Ross makes no explicit predictions as to whether this is the only setting in which lexical metatypy can occur, he describes it as a process that may occur in bilingual situations where a small emblematic language is modeled on a larger intergroup language (see Section 3.1.4.2). This social setting should be distinguished from that of the more frequent transfer of full lexical copies, which may enter a language in situations of casual contact and low degree of bilingualism among recipient language speakers. Thomason and Kaufman (1988: 77) predict that “with a minimum of cultural pressure we expect only lexical borrowing, and then only in non-basic vocabulary”, which they define as culture-specific content words that are typically copied for cultural and/or functional reasons. They add that this type of transfer often happens in socially dominant languages that copy lexical items from numerically smaller groups. Winford (2010: 177) emphasises the role of ‘need and prestige’ for the occurrence of lexical copying, but does not mention any restrictions on the transfer of semantic structures.

4.3 METHODOLOGY AND CATEGORIES OF CLASSIFICATION

4.3.1 METHODOLOGY

As was described in detail in Section 1.2.2, the lexical comparison between Dolgan and Sakha was carried out on the basis of the set of 1,500 meanings from the Loanword Typology list. Although the purpose of the Loanword Typology project was narrower than the current aim to investigate lexical differences in general, the 1,500 item meaning list provides a useful starting point for comparing lexica in a structured way. This section provides the methodological background as well as the analytical framework that was used for the classification of lexical differences between Dolgan and Sakha.

While the registration of differences between Dolgan and Sakha is the first requirement for the establishment of lexical change, the identification alone does not provide any information on the direction of change. In order to arrive at a diachronic ‘change’ from the synchronic state implied by the word ‘difference’, it is necessary to identify a direction and an initiator of change. To achieve this, the attested differences between the Turkic languages Dolgan and Sakha were contrasted on the one hand with the linguistic encoding of these concepts in other Turkic languages and with Tungusic languages (in particular the contact language Evenki) on the other. This applies to differences in linguistic forms as well as to differences in their semantics. Through such a comparison with genealogically related languages on the one hand, and non-related but geographically neighbouring languages on the other it is possible to uncover typical semantic groupings within each language family, as well as deviations from these patterns, which in turn is crucial for the establishment of the direction of change (see also Section 1.2.3).

The general working hypotheses that I adhered to are the following:

- 1 a. If for a certain difference Sakha shows a feature which is typical for Turkic languages, but which is lacking in Dolgan, then it is more likely that the difference reflects a change in Dolgan.
- b. If for a certain feature Dolgan conforms to a general Turkic pattern and Sakha does not, the innovation is more likely to have occurred in Sakha.
- 2 a. If for a certain difference Dolgan is similar to Evenki, and to Tungusic languages in general, contact with these languages could be an explanation for the change in Dolgan.
- b. If the similarity exists just with Evenki and not with other Tungusic

languages, the possibility of contact-induced change in Evenki should be considered.

In practice this involved the following procedure. For each concept for which Dolgan differs from Sakha, a detailed comparison was made of the semantics of the lexical item in both languages based on data from fieldwork (from elicitation and narratives), complemented by extensive dictionaries for Dolgan (Stachowski 1993, 1998) Sakha (Sleptsov 1972, Pekarskij [1907 - 1930] 1958-1959, Voronkin 1995, 1999) and comparative dictionaries and grammars for Turkic (Sevortyan 1974, Erdal 1998, Tenishev 2001). Comparison with other Turkic languages then showed which language follows the general Turkic semantic pattern, and which language deviates from it. For example, a shared semantic grouping for Sakha and other Turkic languages versus a broader meaning in Dolgan could point to an extension of meaning in Dolgan. Conversely, an identical semantic pattern in Dolgan and other Turkic languages versus a different pattern in Sakha could reflect a change in Sakha. The hypothetical scenario where both Dolgan and Sakha differ from other Turkic languages in different ways was not attested and will therefore not be elaborated further. Comparison with Evenki (Boldyrev 1994, Myreeva 2004, Vasilievich 1968) and other Tungusic languages (Doerfer 2004, Benzing 1956), should then clarify whether contact with this language could explain the attested difference. For each concept it was investigated: a) whether the innovative pattern in Dolgan matches the pattern in Evenki; and b) whether Evenki displays a division of conceptual space found more generally within Tungusic, or whether it is an outlier within the family. The latter is important because if Evenki differs from other Tungusic languages, this could reflect a change in Evenki changed due to contact with Turkic languages. Taking all this information together: if Sakha follows the Turkic pattern and Evenki the Tungusic pattern, while the semantic organisation of Dolgan is different from Sakha in a way that matches Evenki, then there is reason to assume semantic transfer from Evenki to Dolgan. If, on the other hand, the different meaning in Dolgan does not match Evenki, the change in Dolgan could have developed as a result of language-internal motivations, such as language attrition. During this process, specific or infrequent words may fall into disuse, which may lead to the development of polysemy in the remaining lexical items.

4.3.2 CATEGORIES OF CLASSIFICATION

Any comparison can be conceived of as the description of one phenomenon with respect to another and involves a standard of comparison and a comparee. Without wanting to make a priori assumptions about the direction of change, for the purpose of terminology I have taken Sakha as the standard of comparison and have described the differences in Dolgan with respect to Sakha. The attested differences between Dolgan and Sakha were classified into six overarching categories that cover the various ways in which forms or meanings diverge in the two languages. These categories are represented in the first column of Table 4.1 and are further clarified below. For certain purposes, these six categories were further specified in the categories in the second column.

Table 4.1: Categories for classification of lexical differences between Dolgan and Sakha

Category	Subcategory
REPLACEMENT	- Russian copy - Evenki copy - Unknown
SEMANTIC CHANGE	- Broader - Narrower - Shift
CHANGE IN FORM	- Phonetic change - Morphological change
DESCRIPTIVE	- Descriptive phrase - Calque
ABSENCE	
?	

REPLACEMENT: for a concept, Dolgan and Sakha use different lexical forms and the form in Dolgan cannot be traced back to Sakha (or Turkic) origin. This category covers copies from Russian and Evenki, but also a few lexical items of unclear origin that could be copies from other languages or language-internal innovations in Dolgan. It is important to note that under the label ‘Russian copy’ two kinds of replacements are subsumed. On the one hand, this label covers cases where for a certain concept Sakha uses a Sakha word and Dolgan uses a word from Russian; on the other hand, it is applied to cases where the Russian copy in Dolgan is different

from the Russian copy that is used in Sakha. Russian copies shared by Dolgan and Sakha are not counted as differences between the two languages, and so are not considered here.

SEMANTIC CHANGE: this label is applied to cases where “[o]nly the meaning of a lexeme changes while other aspects of the item remain constant” (Wilkins 1996: 268). That is, a difference is classified as semantic change if for a concept Dolgan uses a lexical form that is also used in Sakha, but with a different meaning. The difference is classified as ‘broader’ when the Dolgan meaning covers a wider semantic space than it does in Sakha, as ‘narrower’ when it covers less, and as ‘shift’ when the meanings in Dolgan and Sakha do not overlap (anymore). At this point of the analysis, the terminology reflects a synchronic state and no claims are made with respect to the direction of change. Therefore ‘broader’ could mean semantic extension of a lexical item in Dolgan, but could equally well be due to semantic specification or narrowing in Sakha. It simply indicates that the synchronic meaning in Dolgan is broader than that in Sakha.

CHANGE IN FORM: if for a concept Dolgan and Sakha use lexical items with a clearly identifiable common root, but with a difference in phonetic or morphological shape. Differences in phonology are not included, since the phonological systems of Dolgan and Sakha are identical (but see Section 9.3.1.1 for a discussion of differences in the distribution of allophones).

DESCRIPTIVE: if for a concept Dolgan uses a descriptive phrase where Sakha uses a single lexical item, or if Dolgan uses a descriptive phrase different from the one used in Sakha. It is called a *calque* when the structure of the descriptive phrase is based on a clearly identifiable source language e.g. Russian.

ABSENCE: if a lexical item is absent in Dolgan while it is present in Sakha, most probably due to the fact that the concept has no relevance in the Dolgan culture.

?: if there is too little information available about the difference to classify it into any of the above categories. This can be because the words are not recorded in the available dictionaries or because there is uncertainty about the correctness of the word form.

In the current chapter, only **REPLACEMENT** and **SEMANTIC CHANGE** will be discussed in detail. Changes in morphological form is analysed more extensively in Chapter 5, since some of them are not restricted to a few lexical items, but they are part of a more fundamental structural difference between Dolgan and Sakha that has come about through the process of regularisation. The categories **DESCRIPTIVE** and **ABSENCE** will be only marginally discussed because differences of these types

seem to be the result of cultural and geographical differences rather than of contact between languages. Denotations for cultural items often disappear from the lexicon as a result the adoption of a different lifestyle due to migrations, colonisation, or other events that may lead to changes in culture. Of course, such concepts can still be referred to by descriptive phrases, which may later be lexicalised to varying degrees. These changes I classify as descriptive. Since it is hard to determine the degree to which such descriptive phrases are lexicalised, and thus whether they belong to the lexicon or are impromptu collocations, I chose not to attribute too much significance to their occurrence for the current purpose.

4.4 RESULTS

4.4.1 QUANTITATIVE ANALYSIS

The quantitative analysis of the lexical differences between Dolgan and Sakha is carried out from two perspectives, the first being the onomasiological approach. In the onomasiological approach (or the approach of naming, see Blank (2003)), the idea is to look how meanings are encoded in language. This means that for each of the 1,500 meanings from the Loanword Typology list the linguistic encoding a comparison in Sakha and in Dolgan is compared. If the encoding is identical across the two languages, there is no reason to conduct further analysis. However, if there is any dissimilarity, the meaning is included in the list of differences, regardless of whether the difference is a replacement, a semantic change, a change in form, a descriptive phrase, or that it is absent. Since the 1,500 meaning list is taken as a point of departure, each meaning can be counted as a difference only once, and so this total number serves as a basis for the onomasiological analysis, as is reflected in Table 4.2.

This is different for the second perspective, that of type of difference. In this approach the aim is to see how the different types of difference (replacement, semantic change, change in form, absence) are distributed over the semantic fields, in particular for the fields of ‘the body’ and ‘kinship’. Since consultants sometimes gave more than one lexical form for a particular meaning (i.e. near synonyms) more than one type of difference can be associated with a particular meaning. Since in this part of the analysis I am interested in the frequency of types of differences in semantic domains, all responses were then taken into account.

This is important to keep in mind, since it explains why there is a higher number of differences (776 to be precise) in this approach, than from the onomasiological perspective (602). To illustrate this with an example, the meaning GRANDDAUGHTER is expressed by the form *sien* in Sakha. In Dolgan *sien* is not used at all and instead this meaning can be expressed by *oyo-m ki:h-a* [child-POSS.1SG girl-POSS.3SG], with the literal meaning ‘daughter of my child’, or with the Russian word *vnučka*. Both possibilities are considered equally common among Dolgan speakers and are therefore included in the list. Thus, from an onomasiological perspective, this case provides one difference, namely for the meaning of granddaughter. However, from the perspective of type of difference, this example counts as two differences, the one being of the type descriptive, and the other Russian copy.

4.4.1.1 ONOMASIOLOGICAL PERSPECTIVE

Table 4.2 shows the differences in the encoding of the 1,500 meanings across Dolgan and Sakha per semantic field. It shows the total number of meanings included the semantic field, the number of meanings that is encoded differently in the two languages, and the relative percentage of these differences with respect to the total number of meanings in the semantic field. For example, in the semantic field ‘emotions and values’ 27 out of 48 meanings (56.3%) show a difference in encoding when Dolgan and Sakha are compared, and its position at the top of the table suggests that this field shows most internal diversity. The semantic fields are ranked according to the percentage of differences in descending order.

Table 4.2: Number and percentage of differences in encoding of 1,500 meanings between Dolgan and Sakha

Semantic field	No. of meanings in sem. field	No. of different encodings	% of total no. of meanings in semantic field
Emotions and values	48	27	56.3
Social and political relations	36	20	55.6
Law	26	14	53.8
The house	49	25	51
The body	158	78	49.4
Clothing and grooming	60	28	46.7

Warfare and hunting	41	19	46.3
Agriculture and vegetation	74	34	45.9
Animals	116	53	45.7
Sense perception	49	22	44.9
Kinship	85	37	43.5
Basic actions and technology	79	34	43
Religion and belief	26	10	38.5
The physical world	76	30	39.5
Motion	82	29	35.4
Possession	46	16	34.8
Spatial relations	76	24	31.5
Cognition	62	20	32.3
Time	57	19	33.3
Food and drink	97	32	33
Speech and language	42	13	31
Quantity	40	8	20
Modern world	57	9	15.8
Miscellaneous function words	18	1	5.6
Total meanings	1500	602	40.1

As can be seen from this overview, 602 (40.1%) of a total of 1,500 meanings are expressed differently in Dolgan and Sakha. This seems an unexpectedly high number, considering the fact that the languages are so closely related to each other and have often been described as dialects of the same language. However, this percentage includes all types of difference mentioned above, including phonetic differences. A detailed discussion of the types of differences is presented in Section 4.4.1.2.

The five semantic fields with the highest percentage of differences are ‘emotions and values’ (56.3%), ‘social and political relations’ (55.6%), ‘law’ (53.8%), ‘the house’ (51%) and ‘the body’ (49.4%). Although a direct parallel with the results from the Loanword Typology project cannot be drawn, it is worth drawing attention to the relatively high ranking of ‘the body’ and ‘kinship’. In the Loanword Typology project, ‘the body’ is ranked third lowest when it comes to the proportion of non-native lexical items in this field, with an average of 14.2% cross-linguistically. In addition, body parts occupy a quarter of the Leipzig-Jakarta list of terms that are most resistant to being transferred. Even though these results are about full copies only and not about lexical differences in general, they indicate that cross-linguistically ‘the body’ is a stable semantic field, where a high degree of

change is not to be expected, and even less so for languages as closely related as Dolgan and Sakha. The same cross-linguistic conservatism holds for kinship terminology. Although this semantic field has not made it into the Leipzig-Jakarta list, except for the concept of CHILD, this semantic field is cross-linguistically relatively resistant to the adoption of non-native lexical items, with an average of only 15% of foreign copies (Haspelmath and Tadmor 2009: 64). Against this background, the 43.5% of differences between Dolgan and Sakha in this semantic domain clearly stand out and require more in-depth investigation. These two semantic fields are discussed in detail in Sections 4.4.2.1.1 and 4.4.2.1.2.

4.4.1.2 TYPES OF DIFFERENCE

Table 4.3 summarises the types of difference introduced in Table 4.1 and their frequency of occurrence in the comparison of Dolgan and Sakha. They are grouped by their overarching categories (first column) and are further specified in subcategories (second column). The third column lists for each subcategory its number of occurrences in the data set, and the last column does the same for the number and percentage of the overarching categories. The categories are listed in decreasing order.

Table 4.3: Types of difference: frequency of occurrence

Category	Type of difference	No. of instances	Total for category	
			No.	%
SEMANTIC CHANGE	Broader	332	350	45.1%
	Narrower	14		
	Shift	4		
REPLACEMENT	Russian copy	79	129	16.6%
	Evenki copy	29		
	Unknown	21		
FORM	Morphological change	41	121	15.6%
	Phonetic change	80		
DESCRIPTION	Descriptive	94	95	12.3%
	Calque	1		
ABSENCE	Absence	69	69	8.9%
?	?	11	11	1.5%
Total		775	775	100%

The most obvious observation from Table 4.3 is that the overwhelming majority of differences belong to the category of semantic change. The types of difference ‘broader’, ‘narrower’ and ‘shift’ together comprise almost half of all the total number of lexical differences between Dolgan and Sakha (45.1%). Within the category of semantic change, the type ‘broader’ accounts for 94.8% of the instances and is thus the most frequent type of difference, not only within this category, but also within all lexical differences in general (42.8% of all differences). This tendency is not restricted to just one or two semantic fields but seems to be pervasive throughout the entire lexicon. In nineteen of the twenty-four semantic fields, lexical differences between Dolgan and Sakha are dominated by semantic change of the type ‘broader’, and often the occurrence of the second most frequent type of difference is considerably less. The distribution of this type of difference across all semantic fields is presented in Table 4.4. The numbers represent percentages of the total number of differences within the semantic field.

Table 4.4: Types of difference: distribution of subtype ‘broader’ over semantic fields

Semantic field	% of total no. of differences per semantic field
Emotions and values	72.2%
Possession	65%
Quantity	60%
Social and political relations	58.3%
Speech and language	57.1%
Time	53.8%
Cognition	53.5%
Basic actions and technology	53.4%
Motion	52.9%
Kinship	51.6%
Spatial relations	51.5%
Physical world	48.6%
Religion and belief	46.2%
Sense perception	46.1%
Food and drink	40.5%
Animals	37.5%
Clothing and grooming	37.5%
Warfare and hunting	31.8%
Body	30.7%

House	22.2%
Agriculture and vegetation	20%
Law	18.8%
Modern world	10%
Miscellaneous and function words	0

Representing 16.6% (129 instances) of the differences between Dolgan and Sakha, replacement is the second most frequent category. Table 4.5 shows that within this category, 61.2% of the replacements in Dolgan are copies from Russian, including cases where Russian copies replace a Sakha word, as well as cases where they replace a different Russian word; 22.5% are copies from Evenki; and 16.3% are replacements of unclear origin. In relation to the total number of differences these percentages correspond to 10.2% Russian copies, 3.7% copies from Evenki, and 2.7% of unclear origin. For a detailed discussion of the various cases of replacement see Section 4.4.2.2.

Table 4.5: Types of replacement in Dolgan

Category	Type of difference	% of replacements	% of total no. of differences
REPLACEMENT	Russian copy	61.2%	10.2%
	Evenki copy	22.5%	3.7%
	Unknown	16.3%	2.7%
Total		100%	16.6%

Change in form accounts for 15.6% of the differences, 33.9% of which are differences in morphology and 66.1% differences in the phonological realisation of a lexical form. In 12.3% of the differences in form, Dolgan uses a descriptive strategy where Sakha has a single lexical item. This may be due to contact with Evenki, if it is a calque, i.e. in cases where Evenki uses the same descriptive collocation, or it may be part of ongoing language attrition in Dolgan, during which specific lexical items are lost. The 8.9% coded as absence are primarily concepts concerning agriculture, geographical features, and animals that are not present in the environment of the Dolgan people. These concepts have no relevance for their way of subsistence and hence are not lexically expressed in the language. However, absences are observed in other semantic fields such as body parts, which have nothing to do with a difference in culture or geography. In these semantic domains, the absence of lexical items may have to do with a seeming tendency for generalisation in Dolgan, where terms with less specific meanings

take the place of more specific lexical items of Sakha. This impression is strengthened by the large number of differences of the type ‘broader’ and the descriptive strategies for concepts for which Sakha uses a single lexical item. While this impression is only based on the restricted set of concepts of the Loanword Typology list, these data evoke an impression of a tendency in Dolgan towards less specific use of lexical items when compared to Sakha.

4.4.2 QUALITATIVE ANALYSIS

4.4.2.1 SEMANTIC CHANGE

Section 4.4.1.2 showed that the majority of lexical differences between Dolgan and Sakha are semantic changes, and that they were classified more specifically as cases where the lexical item in Dolgan covers a wider conceptual space than the same lexical item does in Sakha. Since the quantitative analysis in Section 4.4.1 showed that the high percentage of differences in the semantic fields of ‘the body’ and ‘kinship’ is cross-linguistically unusual, these fields deserve to be explored more carefully.

4.4.2.1.1 THE BODY

The semantic field of ‘the body’ contains 158 concepts, 101 of which are expressed differently in Dolgan and Sakha. A closer look shows that 33 (32%) are semantic change. The complete overview of all types of difference within this domain can be seen in the Table 4.6 below. From this relatively high number of semantic changes, only three instances could possibly be attributed to contact with Evenki. These are the lexical items for the concepts BEAK, FOOT and BRAIN. With respect to the remaining differences, no evidence for language-external motivation could be found, and they are better explained in terms of language-internal semantic variation or change (see Table 4.10 for an example).

Table 4.6: Distribution of changes in semantic field ‘the body’

Category	Type of change	No.
SEMANTIC CHANGE	Broader	31
	Narrower	1
	Shift	1
REPLACEMENT	Evenki loan	5
	Russian loan	4
	Unknown	7
DESCRIPTIVE		19
CHANGE IN FORM	Morphological change	6
	Phonetic change	18
ABSENCE		7
?		2

In the following examples, which are presented in tables for reasons of clarity, the first column contains the concepts for which semantic change occurred, the next five columns represent the (proto-) languages for which concepts are compared: Proto-Turkic, to show the semantic pattern that is typical for Turkic languages, then Sakha and Dolgan, which are both Turkic languages but map their lexical items differently onto conceptual space. The two rightmost columns show the lexical items and their semantic distribution for Evenki and where possible for Proto-Tungusic to ascertain if Evenki follows the Tungusic semantic pattern. The different shades of grey indicate the shared cognate forms across languages, and the bold borders group together the languages that map lexical items onto conceptual space in a similar way.

Table 4.7: NOSE/BEAK

Concept	Proto-Turkic	Sakha	Dolgan	Evenki	Proto-Tungusic
BEAK	<i>*tum-š-uk</i>	<i>tumus</i>	<i>munnu</i>	<i>oŋokto</i>	<i>*hoŋa+hta</i>
NOSE	<i>*burun</i>	<i>murun</i>	<i>munnu</i>	<i>oŋokto</i>	<i>*hoŋa+hta</i>

Table 4.7 shows that in Sakha the concepts BEAK and NOSE are represented by different lexical items. The same pattern can be reconstructed for Proto-Turkic. In contrast, Evenki and the other Tungusic languages, represented by the Proto-Tungusic reconstruction **hoŋa+hta*, use a single lexical item to express both

concepts. Dolgan shows affiliation to both sides: the lexical form *munnu* is related to Sakha *murun*, but its semantic structure matches the Tungusic model, covering the concepts of NOSE and BEAK. The difference in form between *murun* in Sakha and *munnu* in Dolgan is an interesting topic in itself and will be discussed in Chapter 5.

A similar pattern occurs for the concepts LEG, FOOT and SOLE, as illustrated in Table 4.8.

Table 4.8: LEG/FOOT/SOLE

Concept	Proto-Turkic	Sakha	Dolgan	Evenki	Proto-Tungusic
LEG	*aδak	ataχ	atak	halgan	*palgan
FOOT	*aδak	ataχ/ulluŋaχ	ulluŋ	hagdiki:	*hagdī (-ki)
SOLE	*ultuŋ	ulluŋ	ulluŋ	hagdiki:	*hagdī (-ki)

As can be seen from this table, Dolgan unmistakably uses Turkic lexical items, but the semantic distribution of these items lines up with the semantic patterns in Tungusic languages. As in the previous example, the Turkic and Tungusic languages show different groupings of the concepts LEG, FOOT and SOLE. In Turkic languages a single lexical item reflecting *aδak is used for LEG and FOOT, setting these apart from SOLE. In contrast, Proto-Tungusic and Evenki group together the concepts FOOT and SOLE with a lexical item cognate with *hagdī (-ki), and set it apart from LEG. Dolgan employs the Turkic lexical items *atak* and *ulluŋ*, but their semantic distribution corresponds to that of the Tungusic languages, grouping FOOT and SOLE together as opposed to LEG.

The final example is an instance of semantic shift, where the meaning of the Turkic lexical item for BRAIN in Dolgan has shifted its meaning to denote HEAD.

Table 4.9: BRAIN/HEAD

Concept	Proto-Turkic	Sakha	Dolgan	Evenki	Proto-Tungusic
BRAIN	*bejŋ	meji:	irge	irge	*irgä
HEAD	*töpe	töbö	meni:	dil	*dili

Table 4.9 shows that with respect to the encoding of the concepts BRAIN and HEAD, Dolgan manifests two changes. On the one hand the example displays a

change in semantics of the lexical item *meni:*, which has undergone a semantic shift in Dolgan from denoting BRAIN to meaning HEAD, when compared to Sakha and Proto-Turkic; on the other hand it shows a change in the encoding of the concept BRAIN, resulting in a replacement of the Turkic *meni:* by the Tungusic word *irge* (see also Artemyev 2001: 8). While the replacement itself is palpable, it is impossible to make claims about possible push or pull effects in the sequence of change without historical linguistic data on Dolgan. Was *irge* first copied into Dolgan, leading to the semantic shift of *meni:*? Or had *meni:* in Dolgan first become polysemous for BRAIN and HEAD due to language-internal factors, facilitating the copying of a specific lexical item for BRAIN? Since such historical data for Dolgan are not available, the exact cause of this semantic shift cannot be determined, but from the present use of the Evenki word *irge* in Dolgan it is clear that contact with Evenki played a role in this semantic shift.

For other broader uses of lexical items in Dolgan an explanation in terms of language-internal development is more likely, as is illustrated by the following example featuring the extension of LIVER to BELLY.

Table 4.10: MIDDLE/BELLY/LIVER

Concept	Proto-Turkic	Sakha	Dolgan	Evenki	Proto-Tungusic
INSIDE	? <i>iš</i>	<i>is</i>	<i>is</i>	<i>dō</i>	* <i>dō</i>
BELLY	* <i>karim</i>	<i>is</i>	<i>bïar</i>	<i>hukite</i>	*?
LIVER	* <i>bağır</i>	<i>bïar</i>	<i>bïar</i>	<i>hakin</i>	* <i>pa:</i> - <i>kun</i>

Sakha has one lexical item for INSIDE and BELLY (*is*) and a different word form for LIVER (*bïar*). Dolgan uses the same lexical items, but groups them in a different way; *is* is used for INSIDE only, while *bïar* has the meaning of LIVER and BELLY. However, comparison with both Turkic and Tungusic provides no evidence that this difference must be attributed to contact with Tungusic languages. Evenki has three different lexical items for INSIDE, BELLY and LIVER, so if Dolgan did extend the meaning of *bïar* from LIVER to BELLY, it did not happen according to the Evenki pattern. Similarly, if Sakha extended the meaning of *is* from INSIDE to BELLY, it most likely reflects language-internal change, or potentially contact with a language that was not included into this comparison.

It is remarkable that both Turkic and Tungusic seem to have had specific words for the three concepts and that both Dolgan and Sakha have modified this

pattern, though in different directions. However, cross-linguistically both directions of change are not unusual. According to Wilkins, it is a natural tendency for the meaning of a body part term to “shift to refer to the visible whole of which it is a part” (1981, cited in Wilkins 1996: 273), a pathway that would apply to LIVER → BELLY in Dolgan. The Sakha extension INSIDE → BELLY could be conceived of as an instance of metaphoric change, which is also a common mechanism of semantic change (Geeraerts 2010).

To summarise, out of the 31 semantic extensions in the field of ‘the body’ there are 3 instances that can be plausibly explained in terms of contact between Dolgan and Evenki. However, while such an analysis is very likely after scrutinising the genealogical and geographical patterns through detailed cross-linguistic comparison, the possibility of internally motivated change cannot be ruled out either, especially since the concepts under consideration are semantically closely related. In addition, the examples discussed above are all instances of semantic extension, which is a common language-internal mechanism of semantic change in viable languages, but even more so in situations of language attrition when the use of specific lexical items gradually diminishes and eventually may be lost. When their functions are taken over by the remaining lexical items, these lexical forms become more polysemous and extend their semantic scope. To put it in perspective, a language-internally motivated semantic extension from one part of the body to a spatially contiguous part, as we see in SOLE → FOOT, is cross-linguistically common (Wilkins 1996: 273), whereas the extension NOSE → BEAK is not. On the basis of a cross-linguistic sample, Wilkins (1981, cited in Wilkins 1996: 273-274) formulated five natural tendencies for language-internal semantic change, number four being:

It is a natural tendency for an animal part to shift to refer to a person part (e.g. ‘snout’ → ‘nose’; ‘beak’ → ‘face’). (Wilkins 1996: 274)

The fact that in Dolgan we find the opposite direction of change, which is less likely to occur cross-linguistically, in combination with the fact that the new semantic pattern in Dolgan matches the pattern of Evenki, makes contact between these two languages a very plausible explanation of this change.

4.4.2.1.2 KINSHIP TERMS

Within the semantic field of ‘kinship’, 37 out of 85 (43.5%) concepts are encoded differently in Dolgan and Sakha. In this section I investigate semantic changes in this field, with a particular focus on the concept clusters BROTHER/SISTER, UNCLE/AUNT, MOTHER-IN-LAW/FATHER-IN-LAW, and MAN/HUSBAND-WOMAN/WIFE-FAMILY. For the study of this semantic field, it was not very informative to include Proto-Turkic and Proto-Tungusic reconstructions in the examples, because it turns out that within this time depth the meaning of the proto-forms was often too different from the current meaning to be of help for the reconstruction of a Proto-Turkic or Proto-Tungusic kinship system. However, the terms are referred to in cases where such a reconstruction was possible.

BROTHER/SISTER

Table 4.11 displays the lexical items and their mapping onto conceptual space for the concepts of BROTHER and SISTER in Sakha (and Turkic), Dolgan and Evenki (and Tungusic).

Table 4.11: BROTHER and SISTER in Sakha, Dolgan and Evenki

	Turkic	Sakha	Dolgan	Evenki	Tungusic
OLDER BROTHER OF ♂	*bi:	bi:	ubaj	aki:n	*ak'i
OLDER BROTHER OF ♀	*abaj (?)	ubaj	ubaj	aki:n	*ak'i
OLDER SISTER OF ♂		edzij	edzij	eki:n	*äkä-i
OLDER SISTER OF ♀		ayas	edzij	eki:n	*äkä-i
YOUNGER BROTHER OF ♂	*ini	ini	balis	neku:n	*näkön
YOUNGER BROTHER OF ♀	*jügürči	surus	balis	neku:n	*näkön
YOUNGER SISTER OF ♂	*baldiz	balis	balis	neku:n	*näkön
YOUNGER SISTER OF ♀	*badliz	balis	balis	neku:n	*näkön

Sakha has an elaborate set of terms to refer to siblings with a three-way distinction, depending on: 1) age relative to ego; 2) gender of the sibling; 3) gender of ego, except in the case of YOUNGER SISTER. That is to say, there are different terms for siblings older or younger than ego, for male or female siblings, and for the sibling of a girl or the sibling of a boy, except when the sibling is a younger sister, which in both cases is *balis*. This results in the seven-way distinction for siblings as displayed in Table 4.11. This system matches the general Turkic pattern,

and most of the terms are of Turkic origin as well. Only the terms for ‘older sister’ could be traced back to foreign provenance, of which only *edzij* with certainty. This term is clearly cognate with the Mongolian form *edzi*, whereas *ayas* shows, according to Kałużyński, similarity with both the Old Turkic form *ākā* as well as with Mongolic *egeči* (Kałużyński [1968] 1995: 203). Therefore we cannot unambiguously determine whether this term is cognate with Old Turkic term or with the Mongolic term, or that perhaps knowledge of the Mongolic form influenced the shape of the inherited form *ākā*.

In contrast, Dolgan, makes a terminological distinction for which only a subset of the Turkic criteria is relevant, namely 1) age relative to ego, and 2) gender of the older sibling. Unlike Sakha, the gender of the younger sibling and the gender of ego do not play a role in the terminological distinctions. Table 4.11 clearly shows that Dolgan employs Sakha (and Turkic) terms, but that semantic space is divided up in a different way. More specifically, it shows that the semantic organisation of sibling terms in Dolgan exactly matches that of Evenki, and Tungusic more generally: as in Dolgan, Evenki shows a distinction between older and younger siblings and between male and female siblings that are older than ego. Gender of ego does not play a role, nor does the gender of the younger sibling. This comparison of the three languages (and the families more widely) strongly suggests that the Dolgans use Sakha terms, but have restructured the semantic distribution of these terms on the model of the kinship system of the Evenks.

UNCLE/AUNT

Since the labeling of the kinship relations of AUNT and UNCLE in Sakha and Dolgan is somewhat more complicated than for the previous example, the distribution of terms is given for Sakha, Dolgan and Evenki separately (see Tables 4.12, 4.13, 4.14), building up towards the complete picture. An overview of the entire system, in which the similarities and differences are clearly seen, is given in Table 4.15. In Sakha, aunts and uncles are categorised depending on the gender of ego’s parent (see Table 4.12): father’s brother is called *abaya* and mother’s brother is called *taj*. The relative age of the uncle/aunt to ego’s parent has no influence on the choice of terminology. Although it is not possible to make any claims about a general Turkic system with certainty, it seems to have been similar to the one in Sakha.

Table 4.12: AUNT/UNCLE in Sakha

	Age relative to parent	
	older	younger
BROTHER OF FATHER	<i>abaya</i>	<i>abaya</i>
BROTHER OF MOTHER	<i>ta:j</i>	<i>ta:j</i>
SISTER OF FATHER	<i>edzij</i>	<i>edzij</i>
SISTER OF MOTHER	<i>ta:j edzij</i>	<i>ta:j edzij</i>

According to Tenishev (2001: 296) *ta:j* can be traced back to **taya*, meaning ‘uncle from mother’s side’ (i.e. brother of mother), not mentioning relative age to ego’s parent as a relevant criterion and thus matching Sakha. With respect to *abaya* opinions differ as to whether the term has Turkic (Tenishev 2006: 228) or Mongolic origin (Kałużyński [1962] 1995: 54), but in either case the meaning is ‘uncle from father’s side’ (i.e. brother of father), again not mentioning relative age as a distinguishing feature. The comparison of the category of AUNT between Sakha and other Turkic languages is less straightforward, since for SISTER OF FATHER the Mongolic term *edzij* is used (see previous example), and for SISTER OF MOTHER a combination of a Turkic and Mongolic term. However, despite this deviation from Turkic in the lexical *forms*, the kinship *categories* in Sakha could still match the Turkic pattern, if we assume a symmetrical relation between identification of uncles and aunts. However, this cannot be determined with certainty.

In Dolgan, on the other hand, the gender of ego’s parent is irrelevant but instead the age of the parent’s sibling relative to ego’s parent is the deciding factor.

Table 4.13: AUNT/UNCLE in Dolgan

	Age relative to parent	
	older	younger
BROTHER OF FATHER	<i>ehe</i>	<i>uba/ubaj</i>
BROTHER OF MOTHER	<i>ehe</i>	<i>uba/ubaj</i>
SISTER OF FATHER	<i>ebe</i>	<i>edzij</i>
SISTER OF MOTHER	<i>ebe</i>	<i>edzij</i>

The parent’s brother older than ego’s parent is called *ehe* regardless of whether that is on the father’s or mother’s side. Similarly, parent’s brothers younger than ego’s parent (either mother or father) are called *uba/ubaj*. The same pattern

applies for parent's sisters, labeled *ebe* and *edzij*. An identical pattern is found in Evenki, as is shown in the table below.

Table 4.14: AUNT/UNCLE in Evenki

	Age relative to parent	
	older	younger
BROTHER OF FATHER	<i>ama:ka</i>	<i>aka/aki:n</i>
BROTHER OF MOTHER	<i>ama:ka</i>	<i>aka/aki:n</i>
SISTER OF FATHER	<i>ene:ke</i>	<i>eki:n</i>
SISTER OF MOTHER	<i>ene:ke</i>	<i>eki:n</i>

The lexical forms of Evenki are unrelated to the ones used in Dolgan, but their distribution in conceptual space is exactly the same. Ego's uncles and aunts are labeled differently depending on their relative age to ego's parent, as is the case in Dolgan. Comparison with other Tungusic languages shows that this pattern is common in the entire family: Proto-Tungusic **ama:ka:n* referred to '(mother's) older brother' (Doerfer 2004: 68), and although for *ańaka* no reconstruction is given, the fact that related forms occur in 13 Tungusic languages and dialects is good evidence that this term, and its meaning of '(father's) elder sister' are widespread across the Tungusic family. Now combining the patterns in Sakha, Dolgan and Evenki in one table, we see the following picture.

Table 4.15: UNCLE/AUNT in Sakha, Dolgan and Evenki

	Sakha	Dolgan	Evenki
OLDER BROTHER OF FATHER	<i>abaya</i>	<i>ehe</i>	<i>ama:ka</i>
OLDER BROTHER OF MOTHER	<i>ta:j</i>	<i>ehe</i>	<i>ama:ka</i>
OLDER SISTER OF FATHER	<i>edzij</i>	<i>ebe</i>	<i>ene:ke</i>
OLDER SISTER OF MOTHER	<i>ta:j edzij</i>	<i>ebe</i>	<i>ene:ke</i>
YOUNGER BROTHER OF FATHER	<i>abaya</i>	<i>uba/ubaj</i>	<i>aka/aki:n</i>
YOUNGER BROTHER OF MOTHER	<i>ta:j</i>	<i>uba/ubaj</i>	<i>aka/aki:n</i>
YOUNGER SISTER OF FATHER	<i>edzij</i>	<i>edzij</i>	<i>eki:n</i>
YOUNGER SISTER OF MOTHER	<i>ta:j edzij</i>	<i>edzij</i>	<i>eki:n</i>

While the identical distribution of terms in Dolgan and Evenki itself is suggestive of contact influence, this idea becomes even more appealing as the semantic details of the Dolgan terms are put under the magnifying glass. Table 4.15 shows

that of all the terms for AUNT and UNCLE in Dolgan, only *edzij* is shared with Sakha, be it with only a partially overlapping denotation. In Dolgan, *edzij* denotes a younger sister of ego's parent (regardless of whether father or mother), whereas in Sakha it is the term for the sister of ego's father (regardless of whether older or younger), reflecting once again the importance of relative age to ego's parent in Dolgan versus the importance of gender in Sakha.

While it is not visible in the table above, the other terms used in Dolgan are also shared with Sakha. In Sakha and in Dolgan, *ehe* is used to refer to 'grandfather', *ebe* for 'grandmother' and *ubaj* for 'older brother' (see Table 4.11). However, in Dolgan the semantic coverage of these lexical items is broader than in Sakha, including the meanings of uncle and aunt as well and importantly, the same polysemy is found in Evenki.

Table 4.16 Polysemy of terms used for UNCLE and AUNT in Dolgan and Evenki

	Dolgan	Evenki	Gloss
OLDER BROTHER OF FATHER	<i>ehe</i>	<i>ama:ka</i>	'grandfather'
OLDER BROTHER OF MOTHER	<i>ehe</i>	<i>ama:ka</i>	
OLDER SISTER OF FATHER	<i>ebe</i>	<i>ene:ke</i>	'grandmother'
OLDER SISTER OF MOTHER	<i>ebe</i>	<i>ene:ke</i>	
YOUNGER BROTHER OF FATHER	<i>uba/ubaj</i>	<i>aka/aki:n</i>	'older brother'
YOUNGER BROTHER OF MOTHER	<i>uba/ubaj</i>	<i>aka/aki:n</i>	
YOUNGER SISTER OF FATHER	<i>edzij</i>	<i>eki:n</i>	'older sister'
YOUNGER SISTER OF MOTHER	<i>edzij</i>	<i>eki:n</i>	

In Evenki, the word used for older brother of ego's parent, *ama:ka*, is the same as the word for grandfather, and a younger brother of ego's parent, *aka/aki:n*, also means older brother of ego. The same holds for the terms for sisters of ego's parents: *ene:ke* means older sister of ego's parent and grandmother, while *eki:n* means younger sister of ego's parent and older sister of ego. In Dolgan the pattern is identical. *Ehe* is older brother of ego's parent but also grandfather, and *ebe* is older sister of ego's parent but also grandmother. *Uba/ubaj* is younger brother of ego's parent and older brother of ego, and *edzij* is younger sister of ego's parent and older sister of ego. Thus, the organisation of referential terms for aunt and uncle, as well as the semantic details of the terms chosen for this purpose strongly suggest that the similarities between Dolgan and Evenki are no coincidence but that they have developed as a result of contact between the two populations.

MOTHER-IN-LAW/FATHER-IN-LAW

Table 4.17 displays the terms used to refer to parents-in-law. In Sakha, the terms for parents-in-law are organised both according to the gender of ego and the gender of the parent-in-law. A male ego refers to his parents-in-law as *aya kilin* and *iñe kilin*, a female ego uses the terms *tojon* and *χotun*. While the terms used in Sakha may all be of Turkic origin², it is not clear whether this particular system of reference to parents-in-law is typical for the Turkic language family, since all terms originally had a rather different meaning. For example *kilin* < **qayin* ‘wife’s relatives’ (Tenishev 2001: 309), *tojon* < *tojın* ‘monk’ (Pekarskij [1907 - 1930] 1958-1959: 2706) and *χotun* < **qatyn* ‘wife’ (Tenishev 2001: 296).

Table 4.17: FATHER-IN-LAW/MOTHER-IN-LAW in Sakha, Dolgan and Evenki

	Sakha	Dolgan	Evenki
FATHER-IN-LAW OF ♂	<i>aya kilin</i>	<i>kinnī</i>	<i>etki:</i>
FATHER-IN-LAW OF ♀	<i>tojon</i>	<i>kinnī</i>	<i>etki:</i>
MOTHER IN LAW OF ♂	<i>iñe kilin</i>	<i>iñe kinnī</i>	<i>atki:</i>
MOTHER IN LAW OF ♀	<i>χotun</i>	<i>iñe kinnī</i>	<i>atki:</i>

In Dolgan on the other hand the gender of ego does not play a role. A male and a female ego both use the same terms to refer to their mother- and their father-in-law. The differences in the choice of terms depend solely on the gender of the parent-in-law. As in the previous examples, Evenki uses unrelated lexical items, but their semantic distribution is the same as in Dolgan. In addition, comparative Tungusic etymology shows that this system of reference is deeply rooted in the family: *etki:* < *ekk'in* ‘father in law’ and *atki:* < *atk'i* ‘mother in law’ (Doerfer 2004: 100, 295).

Although linguistic data alone is not sufficient to postulate conclusions with respect to admixture patterns between Dolgans and Evenks in the past, they are an important component within the broader picture including historical, anthropological and genetic evidence. A case in point in this context is the semantic extension of *kilin*, which means ‘parent-in-law of male ego’ in Sakha, and for which the preceding *aya* ‘father’ or *iñe* ‘mother’ specifying the gender of the

² For *tojon* and *qatun* also other origins are suggested.

parent. In Dolgan *kilīn* is extended (after a morphological change leading to the form *kinnī*, see Chapter 5 for details) to denote 'parent-in-law' regardless of the gender of ego, but with a specification of *iīe* 'mother' for 'mother-in-law'. The fact that 'parent-in-law' from the male perspective has been kept and extended in Dolgan could arguably be indicative of a pattern of Evenki women marrying into the Dolgan community. This would be compatible with the tradition of patrilocality in both Dolgan and Evenk communities (Ventsel 2005: 152, personal observation), as well as with the percentage of sharing of mtDNA haplotypes (see Sections 2.6.2, 2.6.4). If an Evenki woman married a Dolgan man and presumably began to learn the Dolgan language, she would have heard most Dolgan speech within her new Dolgan family and from her husband. Given the fact that people normally speak from their own perspective, this means that she would have heard *kilīn* (parents-in-law from the male perspective) more frequently than *χotun* and *tojon* (the Sakha terms for parents-in-law from a female perspective, which may have been used before contact with the Evenks). The husband would have used *aya kilīn* and *iīe kilīn* to refer to her parents (i.e. his parents-in-law), while for her parents-in-law (i.e. his own parents) the husband would have used *kergenner* 'parents'. On hearing *kilīn* being used by her husband for parents-in-law, the Evenki woman, as a second language learner of Dolgan, may have identified this term with the Evenki terms *etki* and *atki*. Through interlingual identification she may have projected the semantic properties of the Evenki terms onto the Turkic word, leading to a generalisation of *kilīn* to denote 'parent-in-law' from the male as well as female point of view.

HUSBAND/WIFE/WOMAN/MARRY

The final example concerns the conceptually related terms for husband, wife and marriage. For this set of concepts Evenki influence is not as compelling as in the previous examples but as will be shown, it could nonetheless help explain the difference in semantic reorganisation between Dolgan and Sakha. It is necessary to point out that this semantic area shows a variety of terms to refer to a single concept, all with their own shades of meaning, especially in Sakha. It is therefore difficult to define a single lexical item as *the* word for husband or for wife. For the same reason, a comparison with other Turkic and Tungusic languages proved not helpful for this example. For the purpose of clarity, only those lexical items that are shared by two or more languages are represented in the table below, but for

the purpose of completeness, the alternatives for HUSBAND and WIFE and FAMILY in Sakha are listed in a footnote.

Table 4.18: HUSBAND/WIFE/FAMILY/MARRY in Sakha, Dolgan and Evenki

	Sakha ³	Dolgan	Evenki
MAN	<i>er</i>	<i>er</i>	<i>edi:</i>
HUSBAND	<i>kergen</i>	<i>er</i>	<i>edi:</i>
WOMAN	<i>d3axtar</i>	<i>d3axtar</i>	<i>asi:</i>
WIFE	<i>kergen</i>	<i>d3axtar</i>	<i>asi:</i>
FAMILY	<i>kergen</i>	<i>kergen</i>	<i>kergen</i>
MARRY	<i>kergennen</i>	<i>erden</i> <i>d3axtardan</i>	<i>edi:le:mi:</i> <i>asi:la:mi:</i>

To start with the first data cell in Table 4.18, *er* in Sakha is used for MAN, whereas in Dolgan it has the added meaning of HUSBAND. To be fair, this is a possible meaning in Sakha as well, but it is not very common⁴. However, Evenki also has a single lexical item to refer to these two concepts, and this model could have reinforced the ‘husband’ aspect of the meaning of *er*. This scenario is supported by the fact that a similar situation applies for WOMAN and WIFE. Dolgan, as well as Evenki, use one lexical item to express both concepts, while for Sakha I have no evidence that *d3axtar* ‘woman’ is used with the meaning of ‘wife’.

Kergen is pervasive in all three languages, but has a more limited meaning of ‘family’ in Dolgan and Evenki, as opposed to ‘family’, ‘wife’ and ‘husband’ in Sakha. Originally the word comes from Mongolic *gergen* ‘wife, married woman’ (Kałużyński [1962] 1995: 156, Lessing 1995: 379a), a meaning that has been kept in Sakha but has over time extended to cover also ‘husband’ and ‘family’. Although not all of the details of this semantic change can be established, the main point here is the observation that this term has the same semantic distribution in Dolgan and Evenki, and that this is different from the semantic distribution in Sakha.

³ Sakha alternatives: HUSBAND: *er* - ‘man’, *oyonrior* - ‘old man’, WIFE: *ojoχ* - ‘woman’ *eme:χsin* - ‘old woman’, FAMILY: *ial* - ‘family’ ‘homestead’ ‘neighbour’ *d3on* - ‘people’ ‘family’.

⁴ In fact the most common way to refer to one’s husband in Sakha is *oyonrior*, at least in the district of Tattaa where I conducted my fieldwork. However, since this term is irrelevant for the current comparison it has been left out.

The meaning of ‘husband’ and ‘wife’ for *kergen* in Sakha has probably facilitated the derivation of the verb ‘to marry’ as well. The verb root

- (1) *kergennen*
 kergen-LA: -(l)n
 spouse-VBLZR-RFL
 ‘to marry’

could literally be translated ‘to spouse oneself’ or in other words ‘to marry’. In Dolgan and Evenki the word for ‘to marry’ is derived from ‘man’/‘husband’ or ‘woman’/‘wife’ depending on the gender of the person who marries. According to Pekarski ([1907-1930] 1958-1959), Sakha has these words as well, but clearly modern Sakha people would use *kergen* first, whereas the Dolgans would not.

4.4.2.2 REPLACEMENT

As mentioned in Section 4.3.2, three categories of replacement were identified in Dolgan, including copies from Russian, copies from Evenki and lexical items of unknown origin. Copies from Russian have entered the language at different stages, whereby a rough division can be made between the pre-Soviet and the Soviet period. As may be recalled from Chapter 2, the nature of the relations between Russians and indigenous peoples was different during each of these stages. This is important to keep in mind because it may have had consequences for the types of change we see in the lexicon as well as in other domains of the Dolgan language today. Russian copies are discussed in Section 4.4.2.2.1; copies from Evenki in 4.4.2.2.2 and lexical items of unknown origin are briefly mentioned in 4.4.2.2.3. An overview of the kinds of replacement, their absolute numbers and percentages is repeated in the table below.

Table 4.19: Replacement in Dolgan

Type of difference	No. of instances	% of replacements	% of total no. of differences
Russian copy	79	61.2%	10.2%
Evenki copy	29	22.5%	3.7%
Unknown	21	16.3%	2.7%
Total	129	100%	16.6%

4.4.2.2.1 RUSSIAN COPIES

On the comparability of the Loanword Typology list for Dolgan and Sakha

Russian copies constitute the largest proportion of replacements in Dolgan. The 79 Russian copies referred to in Table 4.19 make up for 61.2% of all replacements, which corresponds to 10.2% of the total number of lexical differences between Dolgan and Sakha. As was specified earlier, these 79 Russian replacements mean that 79 concepts of the Loanword Typology list are expressed by a lexical item of Russian origin in Dolgan, where Sakha a) uses a non-Russian word or b) uses a Russian word that is different from the one used in Dolgan⁵. However, in practice the overwhelming majority turned out to be of the first type. To quantify this statement, in 68 out of 79 cases (86.1%), the Russian copy in Dolgan replaces a Sakha word of non-Russian origin, leaving only 13.9% for the second scenario. To eliminate confusion, it needs to be emphasized that these 79 differences are counted from an onomasiological perspective. This means that they include cases of polysemy, where a single Russian term was used to express more than one concept in the Loanword Typology list. This in turn means that the absolute number of Russian copies is slightly lower.

While these percentages provide information about the distribution of Russian copies within the subset of the Dolgan lexicon covered by the Loanword Typology list, a thorough comparison of the wordlist for the two languages shows that it is impossible to make any claims about differences in the overall proportion

⁵ This number does not reflect the total number of Russian lexical items. For the current purpose only the differences between Dolgan and Sakha are counted, and therefore meanings for which both languages use the same copy from Russian were not included in the calculations. Therefore, the total percentage of Russian copies is higher.

of Russian copies between the two languages. The lexicon of Dolgan, as well as of Sakha, contains a large set of Russian words, but as it turns out these sets may not, and often do not, overlap. That is, Sakha may employ Russian copies for a subset of concepts of the Loanword Typology list that only marginally overlaps with the subset for which Dolgan uses Russian words. This result is further complicated by the questionable integration of some Russian items into the native lexicon, in other words, whether they are copies that have been accepted by the entire speech community or whether they are only a nonce-borrowing produced as a sign of willingness to fill the slot in the elicitation task. It is in the nature of elicitation data that the results depend heavily on the language proficiency of the language consultants, on the attitude towards (preservation of) the language, or on the reluctance to accept foreign copies as an integral part of the language. While a simple mechanical count would lead to the conclusion that Dolgan and Sakha have exactly the same overall proportion of Russian copies, namely 16.3%, a closer look reveals that this number has come about for both languages in different ways. From my elicited data for Dolgan it becomes clear that my language consultant put in much effort to keep the lexicon as authentically Dolgan as possible. In contrast to the consultant for Sakha, she preferred to leave entries such as ‘beaver’ or ‘oak tree’, which are irrelevant in both languages, empty rather than filling them with Russian words. This desire to fill out every entry in the Loanword Typology list accounts for 31 Russian copies in Sakha, so if they were excluded from the comparison, the percentage in Sakha would be slightly lower than in Dolgan. In a similar vein, my Dolgan informant also avoided Russian words by using Dolgan descriptive expressions, e.g.

- (2) *die* *ürdü-te*
 house top-POSS.3SG
 ‘roof’

instead of Sakha *kiri:sa* ‘roof’, which is based on Russian *kriša*. Also, she would prefer to find archaic words such as *hurbuk* ‘wooden peg’ instead of Sakha *bi:nte*, from Russian *vint* ‘screw’, for concepts which in everyday speech would be expressed through Russian words. While this yields valuable information about archaic words and their cultural connotations, it does not give an entirely realistic representation of the contemporary Dolgan lexicon as used in everyday speech, and, more importantly, makes direct mapping to the Sakha word list problematic. Thus the divergent results for Russian copies in Dolgan and Sakha are caused for a

large part by the optional employment of these items, as well as by the differences in language attitude of my Dolgan consultant and the consultant for Sakha.

With this knowledge in mind, we can say that in many cases where Sakha uses a Russian copy and Dolgan does not, the two realisations may be used interchangeably: Dolgan could use the Russian copy, and Sakha could use a descriptive phrase like Dolgan to express the concept. The reverse, however, is not true: for those entries where Dolgan uses a Russian word and Sakha does not, the Sakha word is often unknown to the Dolgans or it is used with a different meaning. It is these words that constitute the list of differences between Dolgan and Sakha that are discussed below.

Distribution of Russian copies in the Dolgan lexicon

Russian replacements (i.e. copies from Russian that have replaced a Sakha word or a different Russian word) are not limited to particular semantic domains but are pervasive throughout a large part of the Dolgan lexicon. However, they are not equally distributed over the 24 semantic fields, and their proportions vary from 18.3% Russian copies in the most affected semantic domains to 0% in the least affected ones. The five semantic fields with the highest proportion of such Russian replacements are ‘the house’ (18.3%), ‘clothing and grooming’ (13.3%), ‘warfare and hunting’ (12.2%), ‘agriculture and vegetation’ (9.5%) and ‘kinship’ (9.4%).

For most of these semantic domains their high ranking is not unexpected when compared to Sakha, nor from a cross-linguistic perspective. Although the purpose of the Loanword Typology project was different from the current purpose to quantify the difference in copied lexical items between two languages, it may still be curious to view this specific result against the cross-linguistic picture of ‘borrowability’ to get an impression of what is typical and atypical in the distribution of foreign copies across semantic fields. Both in Sakha and cross-linguistically ‘the house’, ‘clothing and grooming’ and ‘agriculture and vegetation’ fall within the top five, while ‘warfare and hunting’ is in positions seven and eight respectively (Tadmor 2009: 64, Pakendorf and Novgorodov 2009: 507). The fact that these semantic fields also appear high on the list of differences between Dolgan and Sakha is an indication that Dolgan took the trend in Sakha a step further: those fields that are generally prone to influence from foreign languages have

experienced even more influence from Russian, either due to more intense contact in the past, or to the increased dominance of Russian in recent times.

The only exception is ‘kinship’, which shows stronger foreign influence in Dolgan both when compared to Sakha and the cross-linguistic average. As mentioned earlier, ‘kinship’ in Dolgan displays the fifth highest proportion of (Russian) replacements, which is remarkably higher than the 19th position this semantic field occupies in Sakha or 21st cross-linguistically. Even though this may seem striking, the differences themselves are not all that significant, since the Russian terms are all used alongside native Turkic lexical items.

The five semantic fields in which the lowest proportion of Russian replacements are found are ‘motion’, ‘religion’, ‘speech and language’, ‘quantity’ and ‘miscellaneous and function words’. Four of these fields do not show any Russian copies at all, only ‘motion’ employs *bolot* for ‘raft’ from Russian *plot*, instead of the Sakha word *a:l*, which was not known to my informant.

The fact that ‘religion’ ranks so low in Dolgan, while it is the highest ranked domain in terms of copying cross-linguistically, is explained by the fact that Dolgan does indeed employ copies from Russian in this semantic field, but since they are identical to the Russian copies that are used in Sakha they do not classify as a difference.

Russian copies replacing Sakha words

The Russian copies in Dolgan that replace Sakha words can be divided into three types: 1) the concept and lexical item are both foreign, while in Sakha the concept and the lexical item are both native; 2) the concept and lexical item in Dolgan are both foreign, while in Sakha the concept is foreign, but the meaning of a native lexical item has been extended to express it; 3) the concept is native but the lexical item is foreign, while in Sakha both concept and lexical item are native.

The first type concerns concepts that were known in the traditional lifestyle of the Sakha, but lost their relevance when groups of Sakha began to move north and adopted a different lifestyle. This mainly applies to semantic fields such as ‘agriculture’, ‘animals’, ‘the house’ and ‘warfare and hunting’. Assuming that this is what happened, it is not surprising that many words connected to these domains have changed in Dolgan. Concepts that belonged to the ‘old’ Sakha lifestyle of cattle breeding and life in the taiga lost their relevance and related lexical items

were lost, while concepts related to the ‘new’ Tungusic lifestyle of reindeer herding in the tundra as well as trade with Russians gained importance and had to be added to the lexicon. Often the lexical items were adopted from the language spoken by the people who introduced the concepts, whether they were Evenks (for the lexicon of reindeer terminology) or Russians (for the lexicon of trade). Thus, Russian copies have been entering the language over a long period of time, starting in the 17th century and continuing today. Interestingly, some of the ‘forgotten’ concepts, especially in the semantic field of agriculture, regained importance during the Soviet regime, when cultural contact with the Russians was particularly intense. The examples below show that Dolgan employs many Russian terms where Sakha has retained the native lexical items. In some cases, the Russian terms have undergone semantic change, as in *ferma*, which has extended from ‘farm’ in Russian to ‘farm’ and ‘stable’ in Dolgan, and *document*, which has extended from ‘document’ in Russian to also mean ‘driver’s licence’, see Table 4.20 and 4.21 below.

Table 4.20: Russian terms where Sakha has native term and native concept

Concept	Sakha	Dolgan	Russian
COW	<i>inaχ</i>	<i>koruoba</i>	<i>korova</i>
STABLE	<i>χoton</i>	<i>ferma</i>	<i>ferma</i> (farm)
PITCHFORK	<i>atird3aχ</i>	<i>vi:la</i>	<i>vila</i>
BUTTER	<i>ari:</i>	<i>sili:be maslata</i>	<i>slivočnoe maslo</i>

The second type is represented in Table 4.21 and exemplifies cases where the concept is foreign to both Dolgan and Sakha, but where Dolgan has adopted a lexical item from Russian and Sakha employs a native word, the meaning of which has expanded to cover the new concept.

Table 4.21: Russian term where Sakha has extended native term and foreign concept

Concept	Sakha	Dolgan	Russian
PLOUGH	<i>χorut</i> (dig)	<i>pahajda:</i>	<i>paxat'</i>
LOCK	<i>χata:hin</i> (bolt)	<i>homuok</i>	<i>zamok</i>
DRIVER'S LICENSE	<i>köŋül</i> (freedom)	<i>dokument</i>	<i>dokument</i>
WALL	<i>erkin</i> (side)	<i>istiene</i>	<i>stena</i>

Aside from copies that came with Russian concepts, Russian lexical items do also occur for concepts that are completely independent of Russian contact and that would have been relevant long before that. This situation was summarised as type three above. Since in these cases the practical need for a Russian word is not so obvious, they could be more illuminating with respect to other potential motivations for the adoption of such foreign lexical items. It may reflect aspects of the possible relationship between Russians and Dolgans in the past and will therefore be discussed in more detail below, despite their small number. From a total of 79 Russian copies in Dolgan that are different from Sakha, six fall into this category.

Table 4.22: Russian form for native concepts

Concept	Sakha	Dolgan	Russian
MOSQUITO	<i>bīrdaχ</i> (<i>kuma:r</i>)	<i>kuma:r</i>	<i>komar</i>
ROOT	<i>silis</i>	<i>kakuora</i>	<i>kokora</i> (hook)
SWAMP	<i>dzebere, kuta</i>	<i>namuluox</i>	<i>navolok</i> (washland)
COLOUR	<i>öŋ</i>	<i>hibiet</i>	<i>tsvet</i>
WEST	<i>arya:</i>	<i>hapad</i>	<i>zapad</i>
DANDRUFF	<i>χoyohō</i>	<i>perxot'</i>	<i>perxot'</i>

The most astonishing result is to find the concept MOSQUITO in this list. Of all living creatures on the Taimyr, these fellows are surely the most prominent and cannot possibly escape the attention of anyone who sets foot on the Peninsula. Importantly, their presence is completely independent of the presence of Russian colonisers. Given the fact that during summer months they dominate the lives of humans and animals it is rather surprising to find that the Dolgan people do not use a native word to refer to this omnipresent phenomenon. Instead of using the Sakha word *bīrdaχ*, they use *kuma:r*, which is clearly cognate with Russian *komar*. According to the dictionary, it is possible to use *kuma:r* in Sakha as well, but during my own visits to the Sakha, *bīrdaχ* was always the default translation and it is used far more frequently.

Silis 'root' exists as *hilis* in Dolgan as well, but according to the Dolgan people I spoke with it has the meaning of 'leaf'. *Kokora* in Russian means 'hook' and can be used in Sakha too to mean 'hook' or 'tree with a hooked root' (Anikin, 2003: 277), while in Dolgan it is taken to mean 'root' in general. So *kakuora* must be classified

as a replacement for *silis* in the meaning of ‘root’, but with a semantic change through polysemy from ‘hook’ to ‘tree with a hooked root’ to ‘root’ in Dolgan.

A confrontation with swamps is also unavoidable when living on the Taimyr, but instead of using Sakha *dzebere* (from Mongolic) ‘swampy mud’, or *kuta* (from Evenki) ‘quagmire’, the Dolgans refer to this phenomenon with *namuluok* from Russian ‘navolok’ meaning a low place on the riverbank where the river leaves its sediments after overflowing. According to Anikin (2003: 387), *namiliaχ* in Sakha means ‘the transfer of a boat by portage from one water body to another’, so it may occur in Sakha, but with a different meaning.

Since these lexical items (can) occur in both Dolgan and Sakha, it is impossible to tell with certainty whether they were copied once into the common ancestor language of Dolgan and Sakha, or whether they were copied at a later stage into both languages independently. According to the first scenario the Russian words would have been copied into a hypothetical common language D/S (Dolgan/Sakha) before groups of Sakha people began to migrate to the north and their language began to diverge. The differences in use and meaning that we see today could then be due to language-internal changes that occurred in the individual languages after their separation. Alternatively, the difference in meaning could be due to the fact that the Russian lexical items were copied into both languages independently and in a different context, and therefore with different connotations.

This category also includes cases where Sakha does not use a Russian word. *Öŋ* ‘colour’ occurs in Sakha, as well as in the Dolgan dictionary by Stachowski (1993: 199), but a preceding asterisk indicates that its occurrence in Dolgan is not certain. The language consultants I worked with did not know the word and produced the Russian *hibiet* instead. DANDRUFF in Dolgan also has no indigenous term but is referred to by the Russian word *perxot*’.

Russian copies for concepts of orientation and cardinal direction are interesting since these concepts seem relevant to everyone, and in particular to nomadic people in a polar desert with very little help of landmarks. Instead of having specific lexical items to express direction, Dolgan uses descriptive terms for all directions, whereas Sakha has a full system of Turkic terms. In addition, the term *hapad* ‘west’ has been copied from Russian *zapad*, which may have to do with the fact that the west is, besides a cardinal direction, also a frequently mentioned socio-political unit.

To summarise, of the three types of Russian copies discussed above, the cases where a Russian copy has been introduced to refer to an already existing concept are most intriguing because there seems to be no objective need for a foreign term. The fact that Russian terminology has pervaded as far into Dolgan as to cover culturally independent concepts such as MOSQUITO or COLOUR could be explained in several ways. Theoretically, it could be a reflection of a numerical dominance of Russian immigrants in the 17th century who introduced these words into the Dolgan-speaking area. The first settlers are said to have interacted with the native population considerably, and to have learned the local language. However, there is no evidence that the number of second language learners of Dolgan was overwhelming. While there were certainly Russian individuals who learned Dolgan as a second language, they are often reported in the literature to have integrated completely with the native population (Dolgikh 1963: 121) and become indistinguishable. While this may be exaggerated, it indicates that they probably knew Dolgan well enough to pick up on words like MOSQUITO or COLOUR. Second, if these Russians spoke Dolgan so well that they practically merged with native speakers, there was no large and distinguishable community of second language speaking Russians, which would be necessary for Russian substrate influence to become established in the Dolgan speech community. Therefore, an explanation in terms of imposition due to numerical dominance of L2 learners of Dolgan is unlikely. More likely the adoption of such words reflects dominance of the Russian language in terms of prestige rather than in number. Maybe the number of Russian people was not large enough to establish substratum effects, but the increasingly dominant status of their language from the 19th century onwards could be a motivation for the introduction of Russian terms in Dolgan, even in the basic lexicon.

Finally, there is the possibility that these copies are relatively recent, and that they are part of the language attrition that is observed as a result of the ongoing shift to Russian. However, this scenario is rather unlikely considering the degree of phonetic integration of these Russian words into the phonological system of Dolgan, which is much stronger in old words than in recently copied ones.

Replacement of Russian words

For 12 concepts both Dolgan and Sakha use a Russian word, but they are different. In most cases this seems to have happened randomly, but in some cases it may reflect linguistic conservativeness of Dolgan when compared to Sakha (see Artemyev 2001a: 9), in that Dolgan uses copies of older dialectal Russian terms, whereas Sakha uses words from modern, literary Russian, as is illustrated in Table 4.23.

Table 4.23: Different Russian copies for the same concept in Dolgan and Sakha

Concept	Sakha	modern Russian	Dolgan	dialectal Russian
BEAUTIFUL	<i>kīrahīabaj</i>	<i>krasivij</i>	<i>baskuoj</i>	<i>boskoj</i>
CALENDAR	<i>χalenda:r</i>	<i>kalendar'</i>	<i>paskal</i>	<i>paskal'</i>

4.4.2.2.2 EVENKI COPIES

General remarks

The number of lexical copies from Evenki is remarkably low given the close relationship between Dolgans and Evenks reported in the literature (see Chapter 2), and the characterisation of Dolgan as ‘Turkic grammar with Evenki lexicon’ or even as a ‘creole’ (Ziker 1998: 102). As was shown in Table 4.19, 3.7% of the overall differences between Dolgan and Sakha are copies from Evenki, which corresponds to 22.5% of all replacements. Statements that the Dolgans originated from different Tungus clans (Popov [1931] 2003: 60) or that they are ‘Yakutized Evenks’ (Dolgikh (1935) cited in Anderson 2000: 86) suggest a very close connection with the Evenks, which could be expected to have had its repercussions on the language. While the current opinion on the origins of the Dolgans is more nuanced, it is undisputed that the Evenks have played an important part in the formation of the Dolgan people, and that there was substantial contact between the two populations (see Chapter 2 for details).

It is important to remember that if a broader range of semantic fields had been included in the analysis, the outcome might have been different. Culturally specific vocabulary, in particular terminology related to e.g. reindeer herding, hide preparation, sleigh riding would have yielded a higher number of copies from

Evenki since it is a known fact that most reindeer terminology in Dolgan was adopted from Evenki. Adoption of culturally specific terminology can thus be an indication of a change in culture, and since these kinds of copies can enter a language also in scenarios of rather superficial contact (Ross 2003: 193) they are not necessarily helpful in the study of population history and possible admixture of peoples. Foreign copies in non-cultural vocabulary occur less easily (Hock & Joseph 1996: 245) and are therefore a more reliable marker of in the study of contact. So even though the number of Evenki copies is higher in Dolgan than in Sakha, the claim that the proportion of Evenki copies constitutes the main difference between Dolgan and Sakha seems to be, with only 3.7% difference, an exaggeration.

Distribution of Evenki copies in the Dolgan lexicon

Since the overall number of replacements from Evenki is not very high, quantitative results for their distribution across semantic fields do not carry much significance. Nevertheless, I consider it worthwhile to give an impressionistic picture on the basis of the available data, from which it appears that the Evenki copies that replace a Sakha word are more restricted in their distribution across the included semantic fields than copies from Russian. The highest percentages occur in the semantic fields of 'the house' (6.1%), 'kinship' (5.9%), 'animals' (3.4%), 'clothing and grooming' (3.3%), 'the body' (3.2%), 'the physical world' (2.6%). Obviously, Evenki has had no influence in domains having to do with modern developments or modern social organisation, such as law, and social and political relations. These spheres are dominated by Russians and the labels for concepts related to these domains were introduced in Russian. More surprising is the fact that Evenki has not left its traces in the domain of 'food and drink', even though this would fall in line with areas such as 'the house', 'clothing and grooming' and 'the body'.

In more than one third of the Evenki replacements (11 out of 29), the Evenki word and the Sakha word are mutually exclusive: the Sakha word does not exist in Dolgan and the Evenki word is not found in Sakha. As can be seen in Table 4.24, for two of the examples this exclusivity exists because Sakha did not seem to have a lexical item for the concept in question.

Table 4.24: Copies from Evenki, Sakha word does not exist

Concept	Sakha	Dolgan	Evenki
SPIDER	oγuj	-	-
	-	ata:ki	ataki:
BUTTERFLY	ürümečči	-	-
	-	lörüö	le:re:
ELK	tajaχ	-	-
	-	tuoki:	to:ki:
YARD	olbuor/telgehe	-	-
	-	nipte	nipte
WEDDING	χolbohu:	-	-
	-	kurum	kurum
NAPE OF NECK	sürün:	-	-
	-	hergi	hergi
CAMP	tühülge	-	-
	-	hara:n	hara:n
VALLEY	χočo	-	-
	-	oγxo	oγxo
PLAIN	sihī	-	-
	-	kitieme	kitieme
IDOL	-	-	-
	-	hemeke:n	hemeke:n
SHEEP	-	-	-
	-	d3ollo	d3ollo

In these cases we can speak of full replacement, since the Sakha word is not remembered by current Dolgan speakers, if it were indeed inherited from Sakha, or does not exist. In the remaining cases the word from Evenki, which sometimes has an adjusted meaning, exists parallel to the word from Sakha, as is shown in the table below, which represents the clearest examples of this type.

Table 4.25: Copies from Evenki, Sakha word exists as well

Concept	Sakha	Dolgan	Evenki
MUSHROOM	<i>tellej</i>	<i>tellej</i>	-
	-	<i>dögömö:xtö</i>	<i>deginmekte</i>
BRAIN	<i>meji:</i>	<i>meji:</i>	-
	-	<i>irge</i>	<i>irge</i>
WHITE	<i>ürüŋ</i>	<i>ürüŋ</i>	-
	-	<i>če:lke:</i>	<i>čelke</i>
THE BOW	<i>ha:</i>	<i>ha:</i>	-
	-	<i>alaŋa</i>	<i>alaŋa</i>
BUTTOCKS	<i>emehe</i>	<i>emehe</i>	-
	-	<i>darama</i>	<i>darama</i>
NAKED	<i>hīgīnńaχ</i>	<i>hīgīnńaχ</i>	-
	-	<i>pelde:ki:n (?)</i>	<i>ńarbaki:n</i>

Tellej is used for mushroom in Sakha as well as in the variety of Dolgan spoken in the Anabar region, which is just across the border of the Taimyr and located in the Sakha Republic. The variety of Dolgan spoken in that area is more similar to Sakha than the Dolgan variety spoken in other areas. In the other Dolgan settlements *dögömö:xtö* is used, which is related to Evenki *deginmekte* (as a third alternative *kuna:χ* can be found, which also exists in Sakha). As far as I am aware, there is no semantic difference between those lexical items. As was mentioned in Section 4.4.2.1.1, *meji:* has undergone a semantic shift in Dolgan. For the concept BRAIN Sakha uses *meji:*, whereas Dolgan uses *irge* from Evenki. However, *meji:* still exists in Dolgan with the meaning of ‘head’. The same holds for the concept WHITE, which is represented in Sakha by *ürüŋ* and in Dolgan by *če:lke:*, as in Evenki. However, *ürüŋ* is also still used in Dolgan, but with the more general meaning of ‘light colour’ or the ‘light colour of reindeer fur’. ‘Bow’ is expressed in Sakha as *ha:*, and in Dolgan as *alaŋa*. *Ha:* exists in Dolgan with the meaning of ‘(unspecified) weapon’. *Emehe* exists in both languages with the meaning of ‘buttocks’ but Dolgan has an additional way of expressing this body part, which comes from Evenki. However, *darama* has undergone a semantic change from ‘crotch’ in Evenki to ‘buttocks’ in Dolgan. Finally, *hīgīnńaχ* exists in both languages with the meaning of ‘naked’, but Dolgan also has the word *pelde:ki:n* which could, according to Voronkin (1995), be related to Evenki *ńarbaki:n* ‘naked’. The difference in meaning between

the two lexical items is that *hiġinńax* in Dolgan is 'naked' but also 'too sparsely dressed', whereas *pelde:ki:n* has the meaning of being completely bare-skinned.

4.4.2.2.3 UNKNOWN ORIGIN

Besides the copies from Evenki or Russian discussed above, there are concepts that are expressed in Dolgan by lexical items that I was not able to trace back to Sakha, Evenki or any of the other neighbouring languages. This may be due to the fact that information in dictionaries is often incomplete, which would plausibly explain the absence of the words in Table 4.26 from the dictionaries at my disposal.

Table 4.26: Lexical items of unknown origin

Concept	Sakha	Dolgan
VAGINA	<i>abas</i>	<i>bökü</i>
PENIS	<i>übüs</i>	<i>öčö:</i>

The Sakha words *abas* and *übüs* (*öbüs* in Dolgan) also exist in Dolgan, but have a rude connotation according to my informants, whereas *bökü* and *öčö:* do not. Alternatively I may not have been able to find, or may not be sure of the related lexical item in other languages due to major changes in phonetic form or meaning, as I could imagine for *χapataj* in Dolgan, meaning 'bald', whereas Sakha uses *tarayaj*, or for *tömüje* for 'finger' where Sakha has *tarbaχ* and the neighbouring Samoyedic language Nganasan has *torija*. Finally, some may just be language-internal innovations part of which could be motivated by onomatopoeic associations (e.g. *titire:*, *čapkahajdas*).

Table 4.27: Lexical items of unknown origin

Concept	Sakha	Dolgan
SHIVER	<i>ilibire:</i>	<i>titire:</i>
ENVY	<i>ĩmsĩ:rĩ:</i>	<i>ordugurgo:</i>
SANDFLY	<i>oŋo:ju</i>	<i>kĩrada:j</i>
SCREW	<i>bi:nte</i>	<i>hurbuk</i> (peg)

SPLASH	<i>ihäx̣tas</i>	<i>čapkahajdas</i>
STORM	<i>sillie</i>	<i>boloho</i>
PIECE	<i>to:roxoj</i>	<i>eltex</i>

4.4.3 SUMMARY

The lexical comparison of Dolgan and Sakha has shown that these languages differ considerably in their lexicon (40.1% of the investigated meanings). However, a closer investigation has shown that a much smaller proportion of these differences can be attributed to contact with speakers of other languages. Since for many differences we cannot determine with certainty whether or not contact played a role in the change, the exact proportion of contact-induced change is hard to determine. To give an estimate, the number of copies from Russian (79) and Evenki (29) added to the probable cases of contact-induced semantic change discussed in this chapter (13) would add up to a proportion of approximately 15% (121:775). Regardless of this relatively low proportion, the character of certain differences is nevertheless indicative of a close relationship between Dolgans and Evenks in the past. This was exemplified by substance copies from Evenki in the domain of non-cultural vocabulary, and more significantly by the restructuring of semantic patterns in the Dolgan system of kinship terms. Other differences were explained through contact with Russians and for a minority of differences the motivations are unclear.

Section 4.4.1.1 showed that from a quantitative point of view the differences between Dolgan and Sakha are not restricted to particular semantic fields. This may indicate that changes are pervasive throughout the entire lexicon, but alternatively this could be due to the rather random allocation of certain concepts in the Loanword Typology list to a particular semantic field. Despite this pervasiveness, certain semantic fields show a higher percentage of differences than others, but the overall picture is a gradual cline rather than a striking pattern, in which only the semantic fields of ‘the body’ and ‘kinship’ stand out in their ranking from a cross-linguistic perspective.

Turning to the nature of these differences, the overwhelming majority was characterised as semantic changes (45.1%) followed by replacements (16.6%) and changes in form (15.6%). The category of semantic change is dominated by the type ‘broader’ (94.8%), where the Dolgan word covers a wider semantic area than

the same word does in Sakha. This type of difference pervades all semantic fields, except function words. Most of these semantic extensions seem to be language-internal developments, and only few can convincingly be argued to have been triggered by contact through comparison with the neighbouring languages. The emerging overall picture of vocabulary with less specific meanings, the relatively frequent use of descriptive phrases where Sakha has a single lexical item, loss of lexical items, and copies from Russian conspires towards the hypothesis that this generalising tendency in Dolgan is a recent development motivated by the ongoing shift to Russian.

However, other differences do point to contact. Within the semantic field 'the body' three semantic changes were argued to be motivated by contact with Evenki. These are the semantic extensions NOSE → BEAK, SOLE → FOOT, and the shift in HEAD → BRAIN. However, for these changes a language-internal motivation cannot be excluded, since the concepts in question are closely related and the direction of change follows cross-linguistic tendencies. The only exception here is the development from NOSE to BEAK, which would be expected to occur in the reverse direction from a cross-linguistic point of view. However, this would not demote contact as a potential explanation of this change. Language-internal and language-external factors are not mutually exclusive, they may have reinforced each other, and the cumulative effect may be reflected in these semantic changes.

Influence from Evenki is more difficult to deny in the explanation of the semantic restructuring of kinship terminology. First, the proportion of instances is higher (61.1%) and second, the restructuring of a kinship system reflects not only a linguistic change, but a more fundamental change in social organisation, which in turn can be explained through close contact and intermarriage between peoples with different social structure. The analysis has shown that Dolgan employs Sakha words for the concepts of BROTHER/SISTER, UNCLE/AUNT, and FATHER-IN-LAW/MOTHER-IN-LAW, but their semantic distribution matches the system of kinship terminology of Evenki.

With respect to foreign copies it appears that Russian copies are more common than copies from Evenki, and that are distributed over more semantic fields. The Russian copies were divided into four types, of which only non-cultural items such as 'mosquito' or 'swamp' could point to a closer relationship between Russians and Dolgans in the past, since they are not a corollary of the introduction of newly introduced concepts of modern society. In the responses to the concepts in the Loanword Typology list there are only six instances of this type, and

although they must certainly not be ignored, it would be overconfident to base bold statements on these few examples. Moreover, most of these Russian copies could be used in Sakha as well, albeit in some cases with a slightly different meaning. Hence, without evidence from historical texts, it is impossible to tell whether this is an instance of a single or two independent copying events. It could be the case that these terms were copied once into the ancestor language of Dolgan and Sakha, after which the meaning diverged in both languages, or it could reflect a situation where the variability in meaning is the result of the fact that the terms were copied into the languages independently. Russian copies of the other types could be due to more intense contact, but could equally well reflect the weaker position of the Dolgan language in the present, and thus be a more recent development.

Copies from Evenki are sparser than Russian copies both in their number and in their distribution. Compared to Sakha, Dolgan shows a higher number of Evenki copies for the investigated semantic fields, but a characterisation of the Dolgan language as Sakha with Evenki lexicon would by no means do justice to the actual facts.

4.5 INTERPRETATION OF RESULTS AND CONCLUSION

In order to formulate hypotheses about the sociolinguistic situation in which the current lexical differences between Dolgan and Sakha developed it is useful to present a schematic overview of the social settings and their expected linguistic outcomes for the different configurations of the Dolgan, Sakha, Evenk and Russian communities in different time frames. Since most contact that is relevant for the purpose of the reconstruction of the Dolgan history took place before Dolgan and Sakha were officially recognised as separate languages, in the schemes I will refer to the common ancestor language as Dolgan/Sakha, abbreviated as D/S.

As may be recalled from Chapter 3, the linguistic outcome of a contact situation is influenced by a complex interplay of linguistic and non-linguistic factors including 1) the structure of the languages in contact; 2) social dominance of the groups in contact; 3) linguistic dominance; 4) attitude and emblematicity. The relationships between the Dolgans and their neighbouring populations differ significantly with respect to these factors, depending on the neighbour in question (in particular the Sakha, Evenks and later the Russians), as well as the time period

during which the contact took place. Therefore it is important to take the various settings into account for the interpretation of the attested lexical differences. Each constellation is specified for the languages in contact, social dominance, linguistic dominance and the linguistic consequences this may have had for Dolgan/Sakha. These factors are considered from the perspective of both communities in contact.

Table 4.28 schematises the contact situation between Dolgans and Evenks until the 20th century, Table 4.29 between Dolgans and Russians during the pre-Soviet period, and Table 4.30 between the same groups after the institution of the Soviet regime. Table 4.31 represents the relation between Dolgans and Russians in most recent times. This division of the Russian contact into three time frames is necessary because the social and linguistic dominance relations were not comparable across these periods and had different linguistic effects on Dolgan.

Table 4.28: Contact situation between speakers of D/S and Evenki

	D/S perspective	Evenki perspective
Social dominance:	D/S	D/S
Linguistic dominance:	D/S	Evenki
Expected effect on Dolgan:	<i>Borrowing:</i> copies of Evenki cultural vocabulary into D/S	<i>Imposition:</i> imposition of semantic structure on D/S through L2 learners

While during the initial period of contact the Evenks may have been socially dominant because they occupied the area before the Turkic-speaking population arrived, these relations quickly changed judging by the establishment of the Khatanga Trading Way, which became associated mainly with the Dolgans, and where D/S became the lingua franca. It is this later configuration, which is represented in Table 4.28. According to this classification, the copies of cultural vocabulary would have entered D/S primarily through a process of borrowing, whereas semantic structures, including changes in kinship terminology, were imposed by Evenki speakers onto D/S. Within the growing community along the Khatanga Trading Way, where Dolgans, Russians and Evenks used to meet and to intermarry, Evenks who learned D/S as a second language may have introduced lexical items and other components from Evenki into their lect of D/S. The fact that such lexical items have become established in the Dolgan language of today could mean that the number of Evenks that learned D/S, and eventually shifted to it, was considerable.

Table 4.29: Pre-Soviet contact between speakers of D/S and Russian

	D/S perspective	Russian perspective
Social dominance:	D/S	D/S
Linguistic dominance:	D/S	Russian
Expected effect on Dolgan:	<i>Borrowing:</i> copies of Russian cultural vocabulary into D/S	

During the pre-Soviet period, the newly arrived speakers of D/S dominated the Taimyr socially as well as linguistically. This holds with respect to other native peoples as well as with regard to Russian settlers. We know that D/S was used as a lingua franca between indigenous people, and historical records report that Russians who came to live there during that time would also learn the language and after a while were ‘indistinguishable’ from the indigenous people. This suggests that the presence of these early Russian settlers did not change the relation of social or linguistic dominance, probably because they were not enough in number and because they were partially dependent on the native population for survival. Therefore Russian substrate effects as a result of imperfect learning are unlikely to have become established in the D/S language. The Russian lexical material that entered D/S during that time can be recognised as labels for unfamiliar cultural items (e.g. food, tools, etc.).

The situation for the Soviet period is different, as shown in Table 4.30.

Table 4.30: Soviet contact situation between speakers of Dolgan and Russian

	Dolgan perspective	Russian perspective
Social dominance:	Russian	Russian
Linguistic dominance:	Dolgan	Russian
Expected effect on Dolgan:	<i>Borrowing:</i> copies of Russian cultural and non-cultural vocabulary into Dolgan due to intense contact, cultural pressure and prestige of Russian	

During this time, Dolgan was already considered a separate language by some scholars, which is why in this table Dolgan is used instead of D/S. From the 1930's onwards, Russian influence, and social dominance, became more and more noticeable in Dolgan society, but until the 1970's most people remained linguistically dominant in Dolgan. However, due to intense contact, cultural pressure and increasing prestige of the Russian language, many cultural as well as non-cultural lexical items were borrowed into the Dolgan language.

The current situation is represented in Table 4.31.

Table 4.31: Current contact situation between speakers of Dolgan and Russian

	Dolgan perspective	Russian perspective
Social dominance:	Russian	Russian
Linguistic dominance:	Russian	Russian
Expected effect on Dolgan:	<i>Imposition:</i> introduction of lexical items from dominant Russian language onto non-dominant Dolgan	

The change in linguistic dominance from Dolgan to Russian took place gradually when people were forced to settle in villages and to go to boarding schools, where the use of Dolgan was forbidden. From then onwards, Russian became more and more socially as well as linguistically dominant, leading to more lexical changes as well as grammatical restructuring that is currently on-going as a result of imposition from their dominant Russian language onto their emblematic, but non-dominant Dolgan (see Chapters 7 and 8 for more changes due to contact with Russian).

CHAPTER 5

REGULARISATION OF PARADIGMS

5.1. INTRODUCTION

Apart from the lexicon, the morphology of Dolgan shows several points of divergence with Sakha. While the overwhelming majority of morphological paradigms is identical in the two languages, a number of differences can be observed, which require closer investigation because they seem to be restricted to the Dolgan-speaking area, and second because they can be subsumed under a common heading of paradigm regularisation, a phenomenon which is not unusual in internally, as well as externally motivated change.

The first phenomenon to be addressed in Section 5.2 is regularisation in the nominal paradigm. It appears that for nouns with a particular phonological structure the forms in Dolgan have a different underlying stem from their cognates in Sakha. More specifically, in Sakha these stems have an irregular declension paradigm, whereas in Dolgan the paradigm has become regular. Counter to previous discussions, in which this difference was assumed to be purely a result of language-internal phonological change, I will argue that this regularisation is the result of a more fundamental cognitive process of reanalysis, motivated, or reinforced by the presence of a substantial number of L2 speakers in the Dolgan-speaking community. The second example concerns the inflectional paradigm of the defective auxiliary verb *e-* ‘to be’ and is described in Section 5.3.

Here the inflectional suffix of the third person plural, which is normally an irregular form within a paradigm of which the endings are otherwise identical to the suffixes of possessive person marking, has been synchronised with the paradigm of possessive person marking, and thus has become more regular. I will argue that this instance of regularisation has occurred on the basis of perceived analogy between the inflectional paradigm of *e-* ‘to be’ and the paradigm of possessive person marking, and that L2 speakers may have played a significant role in the establishment of this change in the speech community. A careful evaluation of language-internal and language-external factors in the development of these changes will be pursued in Section 5.4.

5.2. REGULARISATION OF NOMINAL PARADIGMS

5.2.1 DESCRIPTION OF THE PHENOMENON

In Sakha, the majority of noun stems are inflected in a regular way. They have a transparent agglutinative structure, consisting of a clearly identifiable stem, followed by suffixes of case (ex. 5.1) possession (ex. 5.2), possessive case (ex. 5.3), or predication (ex. 5.4).

SAKHA

- (5.1) **oskuola-ya** *ï:p-pa-ta*
 school-DAT send-NEG-PST.3SG
 ‘She didn’t send me to school.’ (ARR: 41)

- (5.2) *kör-büt-üm,* **oyo-m** *öl-ön* *χa:l-bīt*
 look-PST.PTC-POSS.1SG child-POSS.1SG die-SQ.CV RES-PST.PTC
 ‘I looked, my child had died.’ (ARR: 44)

- (5.3) **inaχ-pītīn** *tut-tu-lar*
 cow-ACC.1PL hold-PST-PL
 ‘They took our cow.’ (ARR: 27)

- (5.4) (...) *ara:s* *buld-u* *bari-tin* *bul-ta:-bīt* **kihi-bin**
 (...) various catch-ACC all-ACC.3SG catch-VBLZR-PST.PTC person-PRED.1SG
 ‘I am a person who hunted all the various animals.’ (AIC: 46)

However, in a small set of Sakha words (the so-called unstable stems) the stem is modified due to morphophonological rules in such a way that the surface form becomes ambiguous to the hearer and opaque with respect to the shape of the underlying stem. For example, the Sakha form *kennitten* ‘from behind’ consists of a stem *kelin* ‘back part’ and a possessive marked ablative suffix with the underlying form *-(t)IttAn*. Due to rules of consonant assimilation and vowel harmony (see Section 5.2.3.3.1) the combination of stem and suffix results in the surface form *kennitten*. In Dolgan, however, the corresponding third person possessive ablative form is *kennititten* (containing an additional syllable *ti*) which consists of a stem *kenni* and the suffix *-(t)IttAn*. The different forms and their underlying morphological structures are presented schematically in Table 5.1.

Table 5.1: Ablative of *kelin* ‘back part’ in Sakha and Dolgan

Language	Ablative	Stem	Translation
Sakha:	<i>kenn-itten</i> <i>kelin-(t)IttAn</i> back.part-ABL.3SG	<i>kelin</i>	‘from behind’
Dolgan:	<i>kenni-titten</i> <i>kenni-(t)IttAn</i> back.part-ABL.3SG	<i>kenni</i>	‘from behind’

Instead of treating *kennitten* as a word with an unstable stem *kelin*, which can only be inferred with the help of complicated rules, it seems that speakers of Dolgan have taken a more straightforward interpretation of the Sakha form *kennitten*. Due to the ambiguity of this surface form, they have taken the ‘mutated’ stem *kenni* as the basis for inflection and have derived the underlying structure directly from this form. This suggests that the inflected form encountered in Sakha has undergone ‘reanalysis’ in Dolgan.

The recognition of such a difference is one thing, but more interesting is the question what could have motivated this change. While seeking to explain this development in Dolgan, particular attention is paid to the question whether the most plausible explanation is found in language-internal processes of change, or whether this phenomenon is better explained by language-external motivations, such as second language learning and language contact. Before addressing these issues in depth, some theoretical background is given on the characteristics of reanalysis in the next section.

5.2.2 REANALYSIS

Reanalysis is an important mechanism of change in syntax and morphology. Aikhenvald defines it as

(...) a historical process whereby a morphosyntactic device acquires a different structure from the one it originally had with little or no change to its surface form or semantics. (Aikhenvald 2006: 30)

Harris and Campbell add that reanalysis “depends upon surface ambiguity or the possibility of more than one analysis.” (Harris and Campbell 1995: 3). One example comes from Udi, a Lezgian language from the East Caucasian language family. For Proto-Lezgian the verb structure is reconstructed as a verb stem preceded by a vowel¹ and a prefix for gender-class according to the following scheme:

(5.5) class marker + vowel + verb stem

Schulze (1982: 148, cited in Harris and Campbell 1995: 67) demonstrates that a number of verbs that had such a structure originally, are nowadays treated by speakers as an unanalysable stem. Thus, a verb like *b+o+q* ‘love, want’, in which *b* is the gender class of ‘other living things’, *o* is the inserted vowel and *q* the original verb stem, would nowadays more accurately be represented as a single unit *boq*. As can be seen from this example, the surface form in both cases is *boq*, but the underlying structure of the form in Proto-Lezgian and in Udi is different, and this corresponds to the definition of reanalysis given above. Not surprisingly, Schulze argues that this development is connected to the fact that Udi is losing the old system of gender-class agreement.

The case of Udi is an example of the loss of morpheme boundaries, but the merging of multiple morphemes into a single unit is not the only way in which reanalysis is manifested. The opposite development is also attested, and speakers can create new boundaries, as happened in the history of the English word *pea*. In the case of *pea* the original singular form was *pease*, and its final *-s* later became interpreted as a plural ending *-s* in analogy with other English plurals ending in *-s* (Lehmann 1992: 223). Thus, reanalysis took place from *pease* > *pea-s* and a new

¹ Harris and Campbell note that the vowel was not originally part of the verb (Harris and Campbell 1995: 66-67).

morpheme boundary was created where there wasn't one before. As third possibility they may shift a boundary to a different position in the morpheme, often motivated by factors such as analogy or iconicity. An example of boundary shift is the French *argent-ier* [silver-AG] 'treasurer' which served as a model for other words such as *bijou-tier* [jewel-AG] 'jeweler', where the *-t* originally belonged to the lexical stem of the noun *argent* and became incorporated in the suffix, leading to a suffix *-tier* (ibid.)

As mentioned before, reanalysis can occur when an alternative analysis of a morpheme (lexical or inflectional) becomes possible, and more plausible, to speakers for various reasons. When both analyses are still available, this results in allomorphy, but eventually it can lead to a permanent change, when the initial allomorph becomes unacceptable to speakers. This process can proceed via the following pathway. Allomorphs can emerge through analogy with other forms in the language (as in the example for *pea*), or through a change in other domains of the language system, in particular phonology. If a phonological change leads to new allomorphs that are ambiguous with respect to their underlying morphological structure, there is the potential for reanalysis to take place (Koch 1996: 237). Whether or not it happens depends on various factors, including economy of processing, frequency of occurrence of the new allomorph (and potentially other forms of the stem) in paradigms as well as in texts, as well as cognitive processes relating to iconicity and markedness.

5.2.3. SAKHA NOMINAL DECLENSION

5.2.3.1. RELATIONAL NOUNS AND REFERENTIAL NOUNS

According to traditional grammatical description, there are two types of nouns in Sakha: independent referential nouns and relational nouns. Although relational nouns do not differ from referential nouns in their inflectional paradigm, the two types do differ from each other in function and context of use, which in turn has consequences for the frequency of occurrence of particular formal properties such as case and possessive marking. To make this more concrete, referential nouns can occupy all main grammatical slots, such as subject and object, and fulfil all basic semantic functions, such as agent, patient and recipient. They denote 'an object or

an objectivised notion², which can occur as an independent unit in the sentence. Relational nouns, on the other hand, are nominal stems that were historically referential nouns, but are used in present day Sakha with grammatical functions and a more figurative meaning. Unlike referential nouns, relational nouns do not occur in basic grammatical functions such as subject and object. They cannot occur as an independent constituent in the sentence and only appear in a dependency relation with other nouns, in particular to specify location, as is illustrated in example 5.6.

- (5.6) *die* *ürdū-te*
 house top-POSS.3SG
 'top of the house'

Since phrases of this type have a schematic structure of NOUN + NOUN-(CASE.)POSS, in which the first noun is the referential noun, and the second one the relational noun, relational nouns almost always occur with possessive and/or case marking (see 5.2.3.2 for more details). As a result, relational nouns hardly ever occur in the unmarked nominative form, which makes it hard for a hearer to determine the underlying stem, particularly if the noun belongs to the category of unstable stems referred to in 5.2.1. This variation in surface form makes the unstable stems more prone to reanalysis than referential nouns, which appear more regularly as a nominative.

5.2.3.2 NOUNS IN PHRASES OF LOCATION

As in many other Turkic languages, dependency relations in Sakha are often expressed by means of the so-called *izafet* construction. This construction, which was copied into Turkic from Persian, expresses a dependency relation between a head noun and a modifier noun by means of an agreement feature (possessive marking) on the head noun. This applies to possessive relations with a literal (ex. 5.7), as well as with a figurative possessive meaning (ex. 5.8), the latter merely establishing a connection between the two nouns, as in phrases of location.

² “Имя существительное – часть речи (особый лексико-грамматический разряд слов), обозначающая предмет или опредмеченное понятие [...]” (Убрятова 1982: 108).

In Sakha, *izafet* constructions consist of a modifier noun in the nominative case, and a possessive marked head noun, which can be marked for case as well. As Sakha is a head final language, the possessive marked head noun occurs in final position in the phrase, and the unmarked modifier noun in initial position, resulting in the schematic structure referred to in the previous section: NOUN+NOUN-(CASE.)POSS.

SAKHA

- (5.7) *elbex aŋar-biṭ Ha:skila:χ-χa ba:l-lar, iŋe-m*
 many half-1PL Saaskylaax-DAT EXIST-PL mother-POSS.1SG
tördü-ler-e
 ancestor-PL-POSS.3SG
 ‘More than half (his relatives) are in Saaskylaakh, the ancestors of my mother.’ (PNL: 13)

- (5.8) *onton bal:an ij-īn otto-tugar köt-ü-t-en*
 then yurt month-GEN middle-DAT.3SG fly-EP-CAUS-SEQ.CV
bardiḷar
 go-PST-PL
 ‘Then in the middle of September they sent him off by plane.’ (XLE: 236)

Phrases of location may express location in space or time, and in order to specify the nature of the relationship between the modifier noun and the head noun, the head noun is specified for case, in particular dative (location), ablative (direction from), and instrumental (mode). Thus, in a locational *izafet* construction, the head noun is always marked for possession to establish the dependency relationship between the head and the modifier, and for case to specify the nature of this relationship. This is illustrated in examples 5.9 and 5.10.

SAKHA

- (5.9) *dzie kenniger χoton ba:r*
 house back.part.DAT.3SG cowshed EXIST
 ‘Behind the house is the cowshed.’ (elicited)
- (5.10) *ip-piṭ bes kennitten (...) kör-ö tur-ar*
 dog-1PL pine back.part.ABL.3SG (...) see-SIM.CV stand-PRS.PTC
 ‘Our dog is looking out from behind the pine tree.’ (elicited)

nominal predicate. Consonant assimilation is a very widespread feature of the language, especially at morpheme boundaries (Stachowski & Menz 1998: 419), and may involve: a) the spread of some phonological features of consonant A to consonant B, leading to more similarity between them but retaining an acoustic boundary; or b) transfer of phonological features across consonants A and B such that these consonants come to share the same set of features, leading to gemination, or doubling, of the consonant. Scenario a) is exemplified in 5.12, where the underlying *-T* in the partitive case suffix *-TA* has become voiced under the influence of the preceding *r* in the stem *ijir* (progressive assimilation).

- (5.12) *ijirde*
ijir-TA
 thread-PART
 '(some) yarn'

In scenario b) two different consonants merge into a single long consonant, whereby the resulting geminated consonant takes the phonological features of the first consonant (progressive assimilation), the second consonant (regressive assimilation), or a subset of features from both (mutual assimilation).

Table 5.2: Assimilation processes in Sakha

Assimilation	Stem	Suffix	Assimilated form	Translation	Assimilation process
Progressive	<i>at</i>	<i>-LAr</i>	<i>at-tar</i>	'horses'	$t \rightarrow l = tt$
Regressive	<i>ba:r</i>	<i>-LAr</i>	<i>ba:l-lar</i>	'they exist'	$r \leftarrow l = ll$
Mutual	<i>at</i>	<i>-Ga</i>	<i>ak-ka</i>	'to the horse'	$t \rightleftharpoons g = kk$

Some scholars propose that every geminated consonant in Sakha are eventually reduced to an assimilation process (Ubryatova 1982: 66). In this view assimilation is indisputable when double consonants appear at morpheme boundaries, and when they occur in the middle of a stem, they must be the result of assimilation between a stem and a suffix, or between two stems, in an earlier stage in the development of the language. In the course of time, they argue, the assimilated form has been reanalysed and become the new stem of the noun e.g. *oloppo*s < *olo*x + *mas* ['seat' + 'wood'] 'chair'. The consonants that can be geminated in Sakha are *p*, *t*, *k*, *l*, *m*, *n*, *ŋ*, *s*, *χ*, *č*. In theory, a geminated consonant can be ambiguous with

respect to the underlying combination of consonants it represents, due to the variety of assimilation processes that occur (e.g. *attar* in Table 5.2 could theoretically be the result of *at+-lar* as well as of *at+-tar*). However, the high token frequency of the assimilated forms, the regularity of their formation, in combination with peoples' exposure to non-assimilated nominative forms, make recognition of the underlying form in most cases an unambiguous task.

5.2.3.3.2. UNSTABLE STEMS

The unstable stems change more significantly under the influence of the suffixes that are attached to them. More specifically, the category of unstable stems discussed here contains bisyllabic nouns with phonological structure (C)V -CV_{high}C. That is, the first syllable has an optional onset, a nucleus that is unspecified for frontness, backness or length and it has no coda. The second syllable of these nouns always has an onset consonant, a nucleus consisting of a high vowel and a coda of one consonant. Examples are the aforementioned stem *kelin* 'back part', as well as *tumus* 'beak', *ürüt* 'top side' and *alın* 'bottom side'. When a suffix is attached to certain noun stems of this type, the high vowel in the final syllable is dropped and the consonants that are consequently adjacent undergo the same assimilation processes as discussed for the regular stems above³.

In cases where the stem ends in a consonant and the added suffix begins with a vowel, it is attached to the formatted stem (which now ends in a consonant cluster), without further modification. For example, *kenne* [*kelin*-(t)A], back.part-POSS.3SG] 'its back part' consists of a stem *kelin* and a third person possessive suffix -(t)A. Since *kelin* is an unstable stem, the high vowel in the final syllable is dropped and the adjacent *l* and *n* undergo assimilation, resulting in a new stem *kenn*. The *t* in the third person possessive suffix is optional and is only inserted if the preceding stem ends in a vowel. Since this is not the case here, only the low vowel *e* (represented by capital *A* in the underlying form according to Turkic tradition) is added to the stem, resulting in a surface form *kenne*.

³ This rule also applies to certain verb forms with this phonological structure (e.g. *hürüt* 'to travel', or the passive on -*ılın*), but they will not be discussed here since with respect to these forms Dolgan does not behave differently from Sakha. This could be due to the fact that for verbs, the stem is identical to the imperative form, and thus occurs in discourse quite frequently.

On the other hand, in cases where the suffix begins with a consonant, an epenthetic high vowel (represented as capital *I*) is inserted between the formatted stem and the suffix. To continue with the same stem as before, the third person plural possessive form of *kelin* is formed by adding the suffix *-LArA* to the stem *kenn*. However, in this case an epenthetic vowel is inserted between the stem and the suffix, possibly to avoid too large consonant clusters and make the inflected form easier to pronounce and to parse. This results in the surface form *kennilere* [*kelin-I-LArA*, back.part-EP-POSS.3PL] ‘their back part’ consisting of *kelin*, an epenthetic vowel *I*, and a third person possessive suffix *-LArA*.

While this may seem many words spent on a small morphophonological detail, it will become clear that this epenthetic vowel has had important consequences for the current shape of unstable stems in Dolgan. Since most suffixes for nominal inflection begin with a consonant, the occurrence of epenthetic vowels with unstable stems in discourse is very high, and as will be shown in Section 5.2.4.3 this may explain for a certain group of reanalysed stems in Dolgan why they end in a high vowel. Since the assimilation process in unstable stems affects the consonant in the middle of the stem as well as those at the stem-suffix boundary, its consequences are more dramatic than in regular stems and determination of the phonological form of the underlying stem becomes less straightforward. However, it is important to note that only a subset of approximately 20 words with this phonological structure behave as unstable stems (see Tables 5.4 and 5.10). Other lexical items such as *kulun* ‘foal’ or *huruk* ‘letter’ have a regular stem and although every account of Sakha mentions this phenomenon as a fact (e.g. Stachowski and Menz 1998: 420), it is not quite clear which factors determine whether a stem is stable or not.

Table 5.3 illustrates the different stages of stem modification in unstable stems for the three situations referred to above: for an unstable stem followed by a suffix starting with a vowel, an unstable stem followed by a suffix starting with a consonant, and a stable stem of the same phonological structure, which does not undergo any change. The first column shows the underlying form of the stem and the suffix, for which optional consonants (as in *-(t)IgAr*) are omitted for the purpose of clarity. The second column shows the form of the word that we would expect to find if the stem were regular and assimilation processes applied as they normally do. Column three displays the crucial characteristic of unstable stems and shows the form of the word after the short high vowel in the last syllable has

been dropped. In column five the word is displayed in its actual shape, after it has undergone the assimilation process mentioned in column four.

Table 5.3: Assimilation processes in unstable stems

1	2	3	4	5	6
Components	Expected form	Vowel drop	Ass. process	Ass. form	Translation
<i>ürüt</i> + <i>IgAr</i> above + DAT.3SG	* <i>ürütüger</i>	* <i>ürtüger</i>	progressive	<i>ürdüger</i>	‘over him’
<i>murun</i> + <i>m</i> nose + POSS.1SG	* <i>murun-u-m</i>	* <i>murn-u-m</i>	regressive	<i>munu-u-m</i>	‘my nose’
<i>kulun</i> + <i>IgAr</i> foal + DAT.3SG	<i>kulunugar</i>	-	-	-	‘to the foal’

Armed with this knowledge about formal properties of unstable stems in Sakha, 5.2.4 explores the differences in form and in use between Dolgan and Sakha. 5.2.4.1 focuses on relational nouns, and 5.2.4.2 does the same for referential nouns.

5.2.4. DOLGAN EQUIVALENTS TO THE SAKHA FORMS

5.2.4.1 RELATIONAL NOUNS

As was foreshadowed in Section 5.2.1, in Dolgan the unstable stems have been reinterpreted in such a way that the assimilated Sakha stem now serves as the root, thus eliminating irregularities due to stem change from the inflectional paradigm. For easy comparison of the forms and their use in Dolgan and Sakha, an additional set of examples is given in 5.13 and 5.14. Here, the form *ürdütünen* [*ürdü-(t)InAn*, upper.part-INST.3SG] ‘on top of’ is the Dolgan alternative to what in Sakha is *ürdünen* [*ürüt-(t)InAn*, upper.part-INST.3SG], clearly displaying the difference in underlying stem.

DOLGAN

- (5.13) *on-tu-ŋ* ***ürdü-tünen*** *ot-tor-u*
 that-DER-POSS.2SG upper.part-INST.3SG grass-PL-ACC
 bīrag-al-lar, *ulaŋan* *buruo* *kel-ien*
 throw-PRS.PTC-PRED.3PL big smoke come-FUT.PTC.ACC.3SG
 ‘On top of that they throw grass, so that there will be much smoke’ (ESB: 71)

SAKHA

- (5.14) *mannik üs-tü:-nen oγoruot ürdünen*
 in.this.way three-DISTR-INST fence upper.part.INST.3SG
kötö hüdʒdʒar uonna ahiliḡar kele turar
 fly-SIM.CV go-PRS.PTC and food come-SIM.CV stand-PRS.PTC
 ‘It jumped over a fence of three bars in this way and came to eat.’
 (PYaI: 109)

While the difference between the inflected forms is obvious enough, the most compelling evidence that the assimilated stem has become the new root in Dolgan is provided by the fact that these stems occur as such in the unmarked nominative case, as in example 5.15. In this example, it is clear that *kenni* (as opposed to the Sakha stem *kelin*) fulfils the function of a referential noun instead of a relational noun, because it is followed itself by a relational noun *diek* ‘side’. *Diek* ‘side’ has been claimed by some scholars to have grammaticalised into an enclitic particle of direction, or even a case form (Ubryatova 1985: 125), while others say it has the status of a noun that can be used independently or as a postposition. The comparative example (5.16) shows that *diek* (or its allomorph *diet*) normally combines with referential nouns like *mas* ‘wood, forest’ in the unmarked nominative case, in other words, with the bare stem of a noun. According to such an analogy, *kenni* must also be analysed as the unmarked nominative case, and thus as the underlying stem of inflected forms such as *kennitiger* (dative) and *kennititten* (ablative).

DOLGAN

- (5.15) *kenni diek kör-dök-püne možet er-bin*
 back.part side look-COND-COND.1SG can.R man-ACC.1SG
gitta arax-s-an kel-iem
 with leave-RECP-SQ.CV come -FUT.1SG
 ‘If I look back I might get divorced from my husband and return.’ (APC: 95)

- (5.16) *Didipte di-en ürex ira:χ, mas diet ba:r*
 Dudypta call -SQ.CV river far wood side EXIST
tüörduon kilometer
 four.ten kilometer.R
 ‘The river Dudypta is far, forty kilometers to the south (lit.: in the direction of the forest).’
 (ANS: 12)

This kind of stem change is not restricted to just the few stems mentioned so far. Table 5.4 gives an overview of other relational nouns that have undergone reanalysis. The first column in the table specifies the language, the second column presents for both languages an inflected form (in this case the third person possessive form of the dative) to illustrate how the unstable stems occur most frequently in actual discourse, the third column shows in bold the underlying stems for both Sakha and Dolgan, and the translation is given in the fourth column.

Table 5.4: Relational nouns in Sakha and their Dolgan equivalents

Language	Dative	Stem	Meaning
Sakha	<i>ürdүger</i> <i>ürüt -(t)IgAr</i>	ürüt	'top side'
Dolgan:	<i>ürdütүger</i> <i>ürdü -(t)IgAr</i>	ürdü	
Sakha:	<i>annıgar</i> <i>alın -(t)IgAr</i>	alın	'bottom side'
Dolgan:	<i>annıtıgar</i> <i>annı -(t)IgAr</i>	annı	
Sakha:	<i>inniger</i> <i>ilin -(t)IgAr</i>	ilin	'front side'
Dolgan:	<i>innitiger</i> <i>inni -(t)IgAr</i>	inni	
Sakha:	<i>kenniger</i> <i>kelin -(t)IgAr</i>	kelin	'back part'
Dolgan:	<i>kennitiger</i> <i>kenni -(t)IgAr</i>	kenni	
Sakha:	<i>onnugar</i> <i>orun -(t)IgAr</i>	orun	'place'
Dolgan:	<i>onmutugar</i> <i>onnu -(t)IgAr</i>	onnu	
Sakha:	<i>ardıgar</i> <i>arıt -(t)IgAr</i>	arıt	'space between'
Dolgan:	<i>ardıtıgar</i> <i>ardı -(t)IgAr</i>	ardı	
Sakha:	<i>attıgar</i> <i>atın -(t)IgAr</i>	atın	'place next to'
Dolgan:	<i>attıtıgar</i> <i>attı -(t)IgAr</i>	attı	

Sakha:	<i>ördüger</i> <i>örüt -(t)IgAr</i>	örüt	'side'
Dolgan:	<i>öttütüger</i> <i>öttü -(t)IgAr</i>	öttü	

In all these cases, the Dolgan stem differs from the Sakha stem in a similar way, suggesting that the change in all these items was brought about by a single underlying process. A possible pathway for this development, in which surface ambiguity of the inflected Sakha form plays a central role, is illustrated in Table 5.5. It shows how the aforementioned Sakha form *ürdünen* [*ürüt-(t)InAn*, top.side-POSS.3SG] may have come to correspond to Dolgan *ürdütünen* [*ürdü-(t)InAn*, upper.part-INST.3SG].

Table 5.5 Possible analyses of *ürdünen* in Sakha and Dolgan

Stem	SAKHA	Possible analyses	DOLGAN	
	INST.3SG		INST.3SG	Stem
<i>ürüt</i>	<i>ürdünen</i>	<i>ürüt -(t)InAn</i>	<i>ürdü-tünen</i>	<i>ürdü</i>
'top side'	top.side .INST.3SG	top.side -INST.3SG		
		<i>ürdü -(l)nAn</i> top.side -INST	top.side - INST.3SG	'top side'

In this table the crucial column is headed 'possible analyses' as it shows that the Sakha surface form *ürdünen* is ambiguous with respect to its underlying structure: on the one hand, a hearer could understand this form as consisting of a stem *ürüt* and an instrumental case suffix, which in theory could be the possessive marked form *-(t)InAn*, as well as the non-possessive instrumental case suffix *-(l)nAn*. However, for a native Sakha speaker, the primary understanding of this form would probably be a possessive marked form for the following reason. Since *ürüt* is a relational noun, it occurs primarily in locational *izafet* constructions, in which the head noun is always marked for possession. Although in this particular form the surface structure is ambiguous with respect to the presence of possessive marking, analogy with other (regular) stems, in which the possessive marking is clearly audible, renders this interpretation for native speakers most likely. This is why the possessive marked underlying form is presented as the first option in the table.

However, from a purely structural, point of view, *ürdünen* is more straightforwardly analysed as a stem *ürdü* and a non-possessive

instrumental *-(l)nAn*. While this interpretation may be less likely for native adult speakers who have in-depth knowledge of the entire linguistic system and its irregularities, such an analysis is easy to imagine for second language learners of Sakha, as well as for young children, who are trying to parse new language forms. After all, the second analysis is much more transparent (since there would be no assimilation involved), economical (since no rules are needed for the inflection of irregular stems) and thus more plausible than the first. Therefore, it is easy to imagine that after having concluded that *ürdū* is a noun stem in Sakha, L2 speakers as well as infants store this form in the mental lexicon. At a later stage, when speakers comprehend that phrases of location involve possessive marked nouns, and in analogy with other possessive marked forms, the suffix *-(t)nAn* is added to the stem *ürdū*, resulting in the innovative form *ürdütünen* that is found in Dolgan today. In small children, such deviant interpretations are typically overruled by the standard usage in the Saha-speaking community. Children will adjust their analysis simply by being exposed to the every-day input of standard Sakha forms or they may be corrected. However, this is not necessarily so for adult second language speakers, and if their number is large enough, or their input of standard Sakha too low, there is a possibility that the deviant interpretation takes root in the L2-speaking community and may even spread among L1-speakers too (see Section 3.1.4 for a detailed review of this scenario).

5.2.4.1.1 VARIATION BETWEEN DOLGAN AND SAKHA STEMS

While the data presented above show that a difference between Dolgan and Sakha in the form of these unstable stems is undeniable, the story would not be complete without mentioning the fact that occasionally the Sakha stems are encountered in Dolgan as well. This variation holds only one way, however: while the Sakha stems are sometimes found in Dolgan, the Dolgan stems are never found in standard Sakha. This is illustrated by the Dolgan examples 5.17 and 5.18, which suggest that the two forms can be used interchangeably. Although the relational noun in 5.17 refers to location in space and location in time in 5.18, other examples from the corpus show that this criterion does not play a role in the choice between *kelin* or *kenni*.

DOLGAN

- (5.17) *onton ol dzaxtar-iŋ bolog-un kennitten*
 then that woman-POSS.2SG balok-ACC.3SG back.part.ABL.3SG
buo gü:le boloχ ba:r
 PRT hall balok EXIST
 ‘Then behind the women’s balok there is the hall balok’ (IMA: 34)

- (5.18) (...) *ol tühe:-bit-im kenni-titten ke kim-iŋ,*
 (...) that dream-PST.PTC-POSS.1SG back.part-ABL.3SG CONTR who-POSS.2SG
ol ös iste-bin buo
 that story hear.SIM.CV-PRED.1SG PRT
 ‘Well after my dream, ehm, I hear that story’ (TJP: 126)

Thus, both stems are acceptable in Dolgan, but they certainly do not occur with the same frequency. The picture is dominated by frequent use of Dolgan stems, supplemented by occasional Sakha stems for a small set of relational nouns, such as *kelin* and *ürüt*. This statement is based on a frequency analysis of relational noun stems in my Dolgan corpus, in which I determined for each noun its overall frequency as well as the number of underlying Sakha stems and Dolgan stems. In this context it is important to note that the underlying stem can only be determined with certainty for a limited number of forms. More precisely, these are the unmarked nominative case (i.e. the stem), and for other cases the third person possessive form. Case forms marked for other persons as well as non-possessive case forms are ambiguous with respect to their underlying stem. As was explained in Section 5.2.3.3.2 it is impossible to determine on the basis of these forms alone whether the high vowel in the middle of the word belongs to the stem (as would be the case when the Dolgan stem is used) or whether it is the epenthetic vowel that is added in Sakha stems between the stem and suffixes that start with a consonant. This is visualised in Table 5.6, in which the possessive paradigm is shown for the dative of *kelin/kenni*, in addition to the nominative form. The forms for which the stem is unambiguous are put in a box. Since the unmarked nominative and the third person singular possessive forms are the only unambiguous forms, the main focus in the discussion of the data will be on those, and the inflectional nouns marked for other persons will play only a marginal role.

Table 5.6: Nominative case and possessive paradigm of dative case *kelin/kenni*

	Surface form		Underlying form	Underlying form
	Sakha	Dolgan	Sakha	Dolgan
NOM.	<i>kelin</i>	<i>kenni</i>		
DAT.1SG		<i>kenniber</i>	<i>kelin -I- BAr</i>	<i>kenni -BAr</i>
DAT.2SG		<i>kenniger</i>	<i>kelin -I- GAr</i>	<i>kenni -GAr</i>
DAT.3SG	<i>kenniger</i>	<i>kennitiger</i>	<i>kelin -(t)IgAr</i>	<i>kenni -(t)IgAr</i>
DAT.1PL		<i>kennibitiger</i>	<i>kelin -I- BItIgAr</i>	<i>kenni -BItIgAr</i>
DAT.2PL		<i>kennigitier</i>	<i>kelin -I- GItIgAr</i>	<i>kenni -GItIgAr</i>
DAT.3PL		<i>kennileriger</i>	<i>kelin -I- LArIgAr</i>	<i>kenni -LArIgAr</i>

The question addressed here only concerns the proportion of Sakha stems with respect to Dolgan stems, therefore spontaneous data as well as elicited data are included in the analysis. The combination of these two sources provides more specific data than would spontaneous speech alone, and does not reduce the reliability of the results, since it is unlikely that text genre influences the choice of noun stem. An overview of both stems in Dolgan is given below.

Table 5.7: Proportion of Sakha stems and Dolgan stems in Dolgan relational nouns

Meaning	Sakha stem	Dolgan stem	Ambiguous	Total
‘back part’	<i>kelin</i> 47.3% (32)	<i>kenni</i> 40.4% (23)	12.3% (7)	62
‘top side’	<i>ürüt</i> 30% (6)	<i>ürdü</i> 60% (12)	10% (2)	20
‘space between’	<i>örüt</i> 28.5% (2)	<i>öttü</i> 57.1% (4)	14.3% (1)	7
‘bottom side’	<i>alın</i> 13.3% (2)	<i>annı</i> 80% (12)	6.7% (1)	15
‘front side’	<i>ilin</i> 0%	<i>inni</i> 88.9% (8)	11.1% (1)	9
‘place next to’	<i>atın</i> 0%	<i>attı</i> 100% (8)	0	8
‘place’	<i>orun</i> 0%	<i>onnu</i> 100% (2)	0	2
Total	42	69	12	123

In Table 5.7 the relational nouns are ranked by occurrence of Sakha stems in decreasing frequency. The most striking observation from this table is that those relational nouns, for which a considerable number of Dolgan and Sakha stems are used, are also the most frequent relational nouns in the Dolgan corpus, with the exception of *örüt*, for which the total number is only seven. In other words, there is a correlation between the presence of a Sakha stem for a relational noun in Dolgan and its overall frequency of occurrence. This claim holds less so for *örüt*, for which the overall number is relatively low but the proportion of Sakha stems quite high, and *alin*, for which the total number is not greatly different from *ürüt* but the proportion of Sakha stems is much lower. However, the correlation for *ürüt* and even more so for *kelin* is so blatant that it is unlikely to be due to chance and therefore requires further investigation. The numbers in this table are based on the frequency of use in Dolgan only, but they do raise the question what the frequency of these relational nouns is in Sakha discourse, and whether the current selection of Sakha stems in Dolgan may be explained by a high frequency of their equivalents in Sakha.

The reason for this hypothesis is the idea that highly frequent items are less likely to undergo a permanent change, since speakers have regular exposure to these forms (Bybee 1991: 72-73). Within this context, frequently used items in Sakha (in this case relational nouns) are more likely to retain their Sakha shape in Dolgan than infrequently used items. The high exposure to these items makes them more likely to be conceived of, and stored in memory as, unanalysable units (like proper postpositions) instead of nouns consisting of a stem and a variable case suffix. Frequency here includes text frequency (the number of times a certain stem occurs in the corpus) as well as paradigmatic frequency (the number of slots a certain stem fulfils in the inflectional paradigm). The data from Sakha that will be presented below suggest that both text frequency and paradigmatic frequency influence the distribution of Sakha stems in Dolgan.

Investigation of the Sakha corpus shows that of all relational nouns *kelin* ‘back part’ stands out as the most frequent one in Sakha oral speech. *Kelin* and its inflected forms with assimilated stems make up for 25.8% of all relational nouns in the Sakha corpus (42 out of 163). To compare, the second most frequent relative noun is *örüt* ‘space between’ with 14.7% (24 instances). The high text frequency of *kelin* in Sakha is mirrored in Dolgan speech, where its equivalent occupies 51.6% of all tokens of relational nouns (33 out of 64). This number is based on only the spontaneous texts for both languages, and includes all relational nouns. Two

inflected forms are particularly frequent in Sakha, namely the ablative form *kennitten* [*kelin-(t)IttAn*, back.part-ABL.3SG] ‘from behind’ and the possessive marked third person singular *kenne* [*kelin-(t)A*, back.part-POSS.3SG] ‘after’. Each of these forms constitutes 38.1% (or 16 in absolute numbers) of all occasions of *kelin* in Sakha. Interestingly, it is exactly these forms, which are found in this shape (i.e. their Sakha shape) in Dolgan.

To support this claim, a comparison of frequencies in Sakha and Dolgan is given below. Table 5.8 presents the frequency distribution in Sakha and Dolgan of the forms with an unambiguous Sakha or Dolgan stem, i.e. the third person possessive forms and the unmarked nominative. Since Dolgan has the option of using both stems, separate columns are created for forms with an underlying Dolgan stem and those with an underlying Sakha stem.

Table 5.8: Frequency distribution of inflectional forms of *kelin/kenni* ‘back part’ in Dolgan and Sakha

Infl. cat.	SAKHA		DOLGAN			
		No.	Dolgan stem	No.	Sakha stem	No.
Nom.	<i>kelin</i> back.part	6	<i>kenni</i> back.part	2		
Nom. 3SG	<i>kenne</i> <i>kelin-(t)A</i> back.part-POSS.3SG	16			<i>kenne</i> <i>kelin-(t)A</i> back.part-POSS.3SG	9
Dat. 3SG	<i>kenniger</i> <i>kelin-(t)IgAr</i> back.part-DAT.3SG	1	<i>kennitiger</i> <i>kenni-(t)IgAr</i> back.part-DAT.3SG	2		
Abl. 3SG	<i>kennitten</i> <i>kelin-(t)IttAn</i> back.part-ABL.3SG	16	<i>kennititten</i> <i>kenni-(t)IttAn</i> back.part-ABL.3SG	3	<i>kennitten</i> <i>kelin-(t)IttAn</i> back.part-ABL.3SG	5
Adj.	<i>kelinji</i> <i>kelin-GI</i> back.part-ADJZR	2				

The table strikingly confirms the idea that Dolgan speakers have only preserved the Sakha version of the forms that occur most frequently in Sakha. This is most clearly illustrated by *kenne*, which is one of the two most frequently occurring forms in Sakha. The table shows that *kenne* occurs in Dolgan rather frequently as well, and most importantly, it exists only in this form. The expected innovative

form **kennite* [*kenni*-(t)A, back.part-POSS.3SG] is not attested in the corpus at all, despite the fact that the nominative form, and thus the underlying stem in Dolgan is *kenni*. This supports the idea that frequently used forms may be stored in the brain as a single unit and thus less prone to change, as is argued for example in Bybee (1991: 77).

Additional evidence, though slightly less stringent, comes from the third person singular ablative form *kennitten*. As for *kenne*, this form with an underlying stem *kelin* is very common in Sakha, and appears in Dolgan in exactly this form as well. However, parallel to this Sakha-based form, Dolgan speakers also use *kennititten*, based on the Dolgan stem *kenni*. This suggests that for some forms two stems are available, which may create confusion regarding the ultimate underlying form for the inflectional paradigm. However, this confusion is unnecessary, if one adheres to the idea that highly frequent forms can become stored as unanalysable units in the mental lexicon. In that case the speaker does not conceive of items such as *kenne* and *kennitten* as consisting of a stem (*kelin*) and suffix *-(t)A* or *-(t)An*, but they would exist as fossilised forms in the mental lexicon. Consequently, it is not necessary to assume an underlying stem *kelin* for these forms, and their existence does not clash with the existence of forms like *kennitiger* and *kennititten*, which are clearly based on an underlying stem *kenni*. These less frequent forms are constructed with the assimilated stem through a productive process of stem + case suffix. This is exemplified by the much less frequent form *kenniger*, of which there is only one instance in Sakha, and which in Dolgan consistently occurs as *kennitiger*, based on the Dolgan stem *kenni* + *-(t)IgAr* [back.part + -POSS.3SG]. Needless to say, it remains impossible to look into the speakers head and leaf through their mental lexicon, but these data suggest that reanalysis of the Sakha stem *kelin* has been completed in Dolgan, resulting in the employment of forms based on the Dolgan stem for the infrequent case forms, while forms based on the Sakha stem (*kenne* and *kennitten*) are lexicalised Sakha ‘islands’, which show the remnants of an earlier stage in the development of the language.

A similar trend, although less pronounced, and less reliable due to the lower number of occurrences, applies to *ürüt* ‘top side’. The left half of Table 5.9 shows the frequency distribution of the relevant inflectional forms of *ürüt* in Sakha, which is clearly dominated by dative and instrumental case forms. The right half of the table displays the occurrence of this relational noun in Dolgan and one can see once again that Sakha stems in Dolgan correspond to the case forms that occur

most frequently in Sakha (dative and instrumental). For *ürdöger* one may argue that it is ambiguous with respect to its underlying structure. After all, *ürdöger* could be analysed as *ürüt-(t)IgAr* [top.side-DAT.3SG] or as *ürdö-GAr* [top.side-DAT.2SG]. However, the discourse context is sufficient to disambiguate this form unequivocally as a dative in the third person possessive and not as a second person, and therefore it can be confidently listed under Sakha stems. Although overall numbers are small (13 for Sakha and 10 for Dolgan), and there is not all that much variation in the Sakha forms either, the results in Table 5.8 at least do not contradict the claim made for *kelin* above, and the more frequently used forms in Sakha are also preserved in Dolgan.

Table 5.9: Frequency distribution of inflectional forms of *ürüt/ürdü* in Dolgan and Sakha

Infl. cat.	SAKHA		DOLGAN			
		No.	Dolgan stem	No	Sakha stem	No.
Nom.			<i>ürdü</i> <i>top.side</i>	3		
Dat. 3SG	<i>ürdöger</i> <i>ürüt-(t)IgAr</i> <i>top.side-DAT.3SG</i>	6	<i>ürdütöger</i> <i>ürdü-(t)IgAr</i> <i>top.side-DAT.3SG</i>	1	<i>ürdöger</i> <i>ürüt-(t)IgAr</i> <i>top.side-DAT.3SG</i>	4
Inst. 3SG	<i>ürdünen</i> <i>ürüt-(t)InAn</i> <i>top.side-INST.3SG</i>	7	<i>ürdütünen</i> <i>ürdü-(t)InAn</i> <i>top.side-INST.3SG</i>	1	<i>ürdünen</i> <i>ürüt-(t)InAn</i> <i>top.side-INST.3SG</i>	1

The data for *örüt* and *alın* are even sparser and therefore they cannot be discussed in great detail. Nevertheless, it is worth mentioning that the Sakha stem *örüt* does occur in Dolgan, despite its relatively low overall occurrence in Dolgan discourse. Importantly, the presence of this Sakha stem needs to be viewed against the knowledge that *örüt* is the second most frequent relational noun in Sakha spontaneous speech (see Table 5.7). In addition, a closer look reveals that the Sakha stem used in Dolgan is a possessive marked dative form *öttöger*, a form which constitutes one third of all the occurrences of this stem in Sakha (29%), and is thus encountered regularly. Although one instance in Dolgan is clearly no basis from which to draw any definite conclusions, it does provide additional support to the general idea that common forms in Sakha are kept in their original form in Dolgan.

Thus, the data from my spoken Dolgan corpus suggest that only the relational nouns *kenni*, *ürdü* and *öttü*, *annı* can occur with both Sakha and Dolgan stems, but

this corpus is not necessarily exhaustive. The overall number of the other relational nouns is rather small and so the absence of Sakha stems may be due to chance. Although a dictionary is not the most reliable source for solving this issue, it may provide supplementary evidence; and indeed, Stachowski's work confirms the presence of both the Dolgan and Sakha stem for the three relational nouns *kenni*, *ürdū* and *anni*, including an additional Sakha stem *ilin* for Dolgan *inni*, which does not occur in my corpus. Stachowski's dictionary and my corpus data agree that for *attī* 'place next to', *onnu* 'place' and *öttū* 'space between' (with one exception), only the Dolgan stem is used. The current variation in use of at least a subset of the relational noun stems suggests that we are witnessing a process of on-going change, which (still) allows both stems to be used, rather than a completed change in the language.

Other factors besides frequency that may condition or restrict the variation in use of noun stems in Dolgan are speaker age or geographical location. Age could affect the distribution of stems if one assumes that an on-going change is most likely to be promoted by the younger generation. In that case one would expect a skewed distribution, with the Dolgan stems occurring more frequently in the younger age groups than in the older generations. However, investigation of the corpus shows that age does not play any role in the distribution of the stem variants. Both stems are used by speakers of all age categories, and without any significant differences in frequency of use.

With respect to geographical location one would expect the people in villages closer to the Sakha border (Syndassko) to use more Sakha stems than the villages further away as a result of regular interaction with Sakha speakers. However, the current data do not explicitly support this expectation. Of the 42 Sakha stems, 18 were produced in Syndassko, the village closest to the Sakha border, 18 in Kheta further to the west and 6 in Volochanka, even more distant from the Sakha border. Thus there is no indication that Sakha stems cluster in the border areas where contact with Sakha is most frequent.

While conclusions with respect to the role of frequency must be drawn with care due to the relatively low number of occurrences in the text corpus, these data provide evidence for a significant role of discourse and paradigmatic frequency in the retention of certain Sakha stems in Dolgan. The skewed distribution of forms in Sakha (i.e. many ablatives and third person possessive forms in the case of *kelin*) may be part of the explanation for the observed correlation between stem type and case form in Dolgan. In other words, it may motivate why exactly these Sakha

forms, and not others, have been preserved in Dolgan, whereas for other forms only the Dolgan stem is used. This retention of Sakha forms does not seem to correlate with other factors such as age or geographical location, and thus stresses the importance of discourse frequency in language variation and change.

5.2.4.2 REFERENTIAL NOUNS

While relational nouns constitute a considerable proportion of unstable stems in Sakha, the observed phenomenon is not restricted to this category of nouns alone. Table 5.10 displays a number of unstable referential noun stems in alphabetical order that have undergone a similar type of morphophonological change as that described for relational nouns. It shows that the noun stems, which have the same phonological structure as the stems discussed above ((C)V-CV_{high}C), have lost their final high vowel and have undergone consonant assimilation just like the relational nouns.

The list in Table 5.10 is not exhaustive, since it only shows the nouns that occur in my corpus. Voronkin (1999) lists a few more lexical items, but since he in fact describes variation in the dialects of Sakha, I do not want to presuppose the existence of all these words in Dolgan without having checked this explicitly. It is striking that more than half of these lexical items consists of body part or kinship terms. This is probably no coincidence, since members of both semantic fields are intrinsically linked to an owner, in a literal sense (body parts) or in a figurative sense (kinship) and therefore are more likely to occur with a possessive suffix than without one. Thus, an unmarked nominative form like Sakha *murun* ‘nose’ is rarely encountered in spontaneous Sakha texts. More common are possessive marked forms like *munnum* [*murun-(l)m*, nose-POSS.1SG], ‘my nose’, *munnuŋ* [*murun-(l)ŋ*, nose-POSS.2SG], ‘your nose’, or a possessive marked case form such as *munnubar* [*murun-l-BAr*, nose-EP-POSS.1SG] ‘on my nose’, which are ambiguous with respect to their underlying stem in a similar fashion as described for relational nouns: *munnum* can be divided up as represented above, where the high vowel in the final syllable is an epenthetic vowel, but from the surface this form could equally well have an underlying structure *munnu-(l)m*, where the high vowel belongs to the stem. Given the higher frequency of the ambiguous possessive marked forms when compared to the non-possessive and non-assimilated forms, and given the opacity of the relation between the surface form and the underlying stem due to the

morphophonological rules, Dolgan speakers may have developed different assumptions with respect to morpheme boundaries in these forms. A possible pathway of this reanalysis, similar to the principle illustrated in Table 5.5 is given in Table 5.10.

The left half displays forms derived from the Sakha stem, the right half shows forms derived from the Dolgan stem and the column ‘possible analyses’ reveals the two underlying morpheme structures that a hearer could infer from the Sakha input. The stem and third person possessive form are displayed for both languages because these are the only forms in which the stem can be unambiguously determined, and therefore they show most clearly the difference between Dolgan and Sakha. The first person possessive form has been included as an example of an ambiguous form, which could have triggered the reanalysis.

Table 5.10: Potential pathway of reanalysis of referential nouns in Sakha and Dolgan

Stem	SAKHA		Possible analyses	DOLGAN			Translation
	Poss.3SG G on -(t)A	Poss.1SG on -(l)m		Poss.1SG on -(l)m	Poss.3SG on -(t)A	Stem	
<i>harin</i>	<i>hanna</i>	<i>hannim</i>	<i>harin-(l)m</i> ----- <i>hanni-(l)m</i>	<i>hannim</i>	<i>hannita</i>	<i>hanni</i>	shoulder
<i>hürün</i>	<i>hünne</i>	<i>hünnüm</i>	<i>hürün-(l)m</i> ----- <i>hünnü-(l)m</i>	<i>hünnüm</i>	<i>hünnüte</i>	<i>hünnü</i>	spinal cord
<i>kilín</i>	<i>kinna</i>	<i>kinnim</i>	<i>kilín-(l)m</i> ----- <i>kinni-(l)m</i>	<i>kinnim</i>	<i>kinnita</i>	<i>kinni</i>	father in law
<i>köyüs</i>	<i>köxsö</i>	<i>köxsüm</i>	<i>köyüs-(l)m</i> ----- <i>köksü-(l)m</i>	<i>köksüm</i>	<i>köksüte</i>	<i>köksü</i>	back
<i>murun</i>	<i>munna</i>	<i>munnum</i>	<i>murun-(l)m</i> ----- <i>munnu-(l)m</i>	<i>munnum</i>	<i>munnuta</i>	<i>munnu</i>	nose
<i>törüt</i>	<i>tördö</i>	<i>tördüm</i>	<i>törüt-(l)m</i> ----- <i>tördü-(l)m</i>	<i>tördüm</i>	<i>tördö</i>	<i>törüt</i>	ancestor, root
				<i>tördüm</i>	<i>tördüte</i>	<i>tördü</i>	clan, root
<i>tumus</i>	<i>tumsa</i>	<i>tumsum</i>	<i>tumus-(l)m</i> ----- <i>tumsu-(l)m</i>	<i>tumsum</i>	<i>tumsa</i>	<i>tumus</i>	cape, island
				<i>tumsum</i>	<i>tumsuta</i>	<i>tumsu</i>	protruding object, top

DOLGAN:

- As can be seen from Table 5.10, for some nouns, such as *törüt* and *tumus*, both the Dolgan and the Sakha stems are used. However, according to the dictionary they have different semantic connotations. For example *tumus*, which has the meaning ‘beak’ in Sakha, is found in Dolgan as *tumus*, with the meaning of ‘cape, island’, and as *tumsu*, with the meaning ‘top’ or ‘protruding object’ more generally. While for *törüt* I have no evidence other than the dictionary, the semantic difference between *tumus* and *tumsu* are confirmed by data from my own corpus as well. However, the details of this semantic specialisation of cognate stems require further research.

The difference in unstable stems between Dolgan and Sakha for relational as well as for referential nouns has been recognised by other scholars. Voronkin observes in his overview of Sakha dialects that “[i]n the northern dialects (more regularly in the northwestern dialects), a particular formation of the possessive forms is observed (about 20 nouns)”⁴. Note that in his account, Dolgan is viewed as a dialect of Sakha, but this point of discussion is not relevant for the treatment of the data here. He continues that there are different interpretations of the phenomenon.

⁴ “В северных говорах (более регулярно в северо-западных) отмечается своеобразное оформление притяжательной формы (около 20 имён).” (Voronkin 1999: 140).

like Sakha *orun* ‘place’ would have developed into Dolgan *onnu* in the following way: *orun* (→ metathesis) *ornu* (→ assimilation) *onnu*. Voronkin argues that this account is more plausible because there is no need to include the mysterious double possessive marking, and because metathesis is rather common in other related languages, e.g. Tatar *boron* → *bornu* ‘nose’. While the process of metathesis itself may be plausible, it leaves unexplained why it would have happened in the first place. The account based on reanalysis that has been hinted at throughout the chapter, and that will be further explored in Section 5.4, builds on cognitive principles instead and gives insight into this development on a deeper level.

Stachowski (e.g. 1993: 144) also refers to the link between the Dolgan and the Sakha stems when he describes the etymology of Dolgan stems in his dictionary. However, he assumes that the Dolgan stems of this type are derived from the third person possessive form in Sakha, for which he assumes a fossilised (‘erstarrt’) possessive suffix of the form *-(t)I*. The high vowel in this suffix should account for the final high vowel we find in modern Dolgan stems. More precisely, Stachowski analyses the Dolgan stems as coming from a Sakha form which looks like NOUN-*-(t)I*. This suffix, which he sometimes calls a ‘third person marker’ and sometimes a third person possessive marker, was attached to the noun stem, and after this suffixation consonant assimilation took place, as in e.g. *murun-u* > *munnu* ‘nose’, *sarīn-i* > *hannī* ‘shoulder’ and *kelin-i* > *kenni* ‘back part’. The fact that Stachowski assumes *-(t)I* as the third person singular possessive suffix is somewhat puzzling, because the current Sakha suffix for this category is unequivocally recognised in grammars as *-(t)A*. In contrast to the form proposed by Stachowski, this suffix contains a low vowel, and so do the forms that carry it, e.g.: *munna*, *hanna*.

There are two possible explanations for why Stachowski takes *-(t)I* as the underlying form, but as I will show, they are both unsatisfying to explain the observed difference between Dolgan and Sakha. On the one hand, Stachowski may have based his analysis on an idea he expressed elsewhere that *-(t)I* can be extracted as a third person singular marker from oblique case forms of the possessive declension in Sakha (Stachowski and Menz 1998: 422). According to this argumentation, possessive marked case suffixes in Sakha can be subdivided into a part that encodes possession and a part that encodes case. However, the form of these parts does not exactly match the forms the possessive suffixes have in their isolated form. For example, the third person possessive dative suffix *-(t)IgAr* is analysed as a third person possessive marker *-(t)I* and a dative case ending *-gAr*,

and an instrumental case as consisting of $-(t)I$ and $-nAn$. To illustrate the motivation for this idea, an overview of all the third person singular case forms is given in Table 5.11. For comparison the case forms for the first person are also included, as well as the non-possessive case forms.

Table 5.11: Possessive and non-possessive case forms for third and first person singular (Stachowski and Menz 1998: 422).

CASE	NON-POSS.	POSSESSIVE					
		POSS. 1SG	Poss. suff.	Case suff.	POSS. 3SG	Poss suff.	Case suff.
Nom	-	<i>-Im</i>	<i>-Im</i>	-	<i>-(t)A</i>	<i>-(t)A</i>	-
Dat	<i>-GA</i>	<i>-BAr</i>	<i>-BA</i>	<i>-r</i>	<i>-(t)IgAr</i>	<i>-(t)I</i>	<i>-gAr</i>
Acc	<i>-(n)I</i>	<i>-BIIn</i>	<i>-BI</i>	<i>-n</i>	<i>-(t)In</i>	<i>-(t)I</i>	<i>-n</i>
Part	<i>-TA</i>	<i>-BIInA</i>	<i>-BI</i>	<i>-nA</i>	<i>-(t)InA</i>	<i>-(t)I</i>	<i>-nA</i>
Abl	<i>-(t)tAn</i>	<i>-BIItAn</i>	<i>-BI</i>	<i>-ttAn</i>	<i>-(t)IttAn</i>	<i>-(t)I</i>	<i>-ttAn</i>
Inst	<i>-(I)nAn</i>	<i>-BIInAn</i>	<i>-BI</i>	<i>-nan</i>	<i>-(t)InAn</i>	<i>-(t)I</i>	<i>-nan</i>
Comit	<i>-LL:n</i>	<i>-BIInA:n</i>	<i>-BI</i>	<i>-nA:n</i>	<i>-(t)InA:n</i>	<i>-(t)I</i>	<i>-nA:n</i>
Comp	<i>-TA:γAr</i>	<i>-BIInA:γAr</i>	<i>-BI</i>	<i>-nA:γar</i>	<i>-(t)InA:γar</i>	<i>-(t)I</i>	<i>-nA:γar</i>

On the basis of this overview, the argumentation for $-(t)I$ as a third person marker seems quite plausible, since it is a constant factor in all cases except the nominative. However, it is questionable whether this analysis is sufficient for a satisfactory analysis of *murun-(t)u > munnu*. One may ask whether speakers actually analyse complex morphemes such as $-(t)IgAr$ or $-(t)InAn$ as consisting of a person and a case component. More likely, these forms are processed, produced and stored as a single unit, which is supported by the phonological reduction of some possessive-marked case forms (e.g. Sakha $-(t)IgAr \rightarrow -Ar$). The high level of analytical skill on the part of the speaker and consciousness of the internal morpheme structures make this explanation unattractive and implausible.

A more realistic explanation of the difference between Sakha *murun* and Dolgan *munnu* is provided by the fact that these forms are typically used in their possessive form in discourse whereby an epenthetic high vowel is inserted between the stem and the possessive (case) suffix. As for the relational nouns, the morphological structure of the resulting surface form is ambiguous for the hearer, and the epenthetic vowel can be analysed as part of the inflection, as is the case in Sakha, or it can be analysed as part of the stem, as is the case in Dolgan. Since this epenthetic vowel appears in every possessive marked case form except for the

third person, the assimilated stem followed by a high vowel has a high text frequency, which makes this form a suitable candidate for reanalysis. To illustrate this, the possessive paradigm for the stem *murun* is displayed in Table 5.12, showing the nominative, accusative and dative case. From this it is clear that only in the third person no epenthetic vowel is inserted. And even of those third person forms, only in the nominative case this results in a low final vowel.

Table 5.12: Possessive inflection for *murun* 'nose' (NOM., ACC., DAT.)

	NOM	ACC	DAT
1SG	<i>munnum</i> <i>murun-I-m</i>	<i>munnubun</i> <i>murun-I-BIn</i>	<i>munnubar</i> <i>murun-I-BAr</i>
2SG	<i>munnun</i> <i>murun-I-n</i>	<i>munnugun</i> <i>murun-I-GIn</i>	<i>munnugar</i> <i>murun-I-GAr</i>
3SG	<i>munna</i> <i>murun-(t)A</i>	<i>munnun</i> <i>murun-(t)In</i>	<i>munnugar</i> <i>murun-(t)IgAr</i>
1PL	<i>munnubut</i> <i>murun-I-BIt</i>	<i>munnubutun</i> <i>murun-I-BItIn</i>	<i>munnubitigar</i> <i>murun-I-BItIgAr</i>
2PL	<i>munnugut</i> <i>murun-I-BIt</i>	<i>munnugutun</i> <i>murun-I-GItIn</i>	<i>munnugutugar</i> <i>murun-I-GItIgAr</i>
3PL	<i>munnnulara</i> <i>murun-I-LArA</i>	<i>munnnularin</i> <i>murun-I-LArIn</i>	<i>munnnularigar</i> <i>murun-I-LArIgAr</i>

Thus, the high paradigmatic and text frequency of the assimilated stem followed by a high vowel may explain why this sequence has been interpreted as a new stem in Dolgan.

On the other hand, Stachowski may have assumed that the vowel in the third person singular possessive suffix was high at some stage in the history of Sakha, and has changed into a low vowel over time. This is not unimaginable, since the third person singular possessive in other Turkic languages often contains a high vowel, e.g. *-(s)I(n)* in Old Turkic and Turkish (Erdal 1998: 142, Csató and Johanson 1998: 212), *-(s)I* in Kirghiz, Middle Kipchak, Azerbaijanian and Turkmen (Kirchner 1998: 347, Berta 1998: 160, Schöning 1998: 252, 264). However, for this to be a satisfactory explanation, the vowel change must have taken place in Sakha only after the Dolgan people diverged from the Sakha and their characteristic Dolgan speech had been established, since the Dolgan stems all contain a high vowel in the final syllable. In other words, the change in Sakha must have happened after the 17th century. Although there is very little data on Sakha that go back to the 17th

or 18th century (the earliest mention of the Dolgan people), old word lists show that the POSS.3SG suffix in Sakha already contained a low vowel in the beginning of the 19th century. Sauer's word list, which was compiled in 1803, contains a number of items, which testify that the vowel in the third person possessive was already low at that time. For example, there is the form *a:t-a*⁵ [*a:t* -(t)A, name-POSS.3SG] 'his name' (Sauer 1803: 318) and *tī:n-a*⁶ [*tī:n* -(t)A, breath-POSS.3SG] 'his breath' (ibid.: 7). An additional problem of such an analysis is the fact that every other third person possessive form in Dolgan ends in a low vowel, just as in modern-day Sakha. If the above scenario were true and the third person possessive vowel in Sakha was lowered only after Dolgan diverged from Sakha, it is hard to explain why Dolgan has not retained a high vowel in other third person singular possessive forms, in particular since Dolgan is assumed to have preserved archaic aspects of Sakha in other domains. These two arguments render an account of the Dolgan stems that is based on a third singular possessive form with a high vowel highly unlikely, even if this final vowel were high in Sakha at an earlier stage.

To summarise, Section 5.2 has shown that relational nouns as well as referential nouns which in Sakha have the phonological structure (C)V-CV_{high}C, often have a different shape in Dolgan, namely (C)VC-CV_{high}. For both relational and referential nouns we have seen that the Sakha stems can also sometimes be used in Dolgan, although the Dolgan stems occur in the majority of cases. In the case of relational nouns it was postulated that this is connected to frequency of use in discourse, where particularly common forms in Sakha (such as *kenne* and *kennitten*) may have become stored as unanalysable units in the speaker's mental lexicon as a result of high input frequency. Some referential noun stems can also occur in two versions (such as *tumus* 'cape'). In this case each stem variant has come to emphasise different aspects of the Sakha meaning, and nowadays they occur in different contexts. The extent to which the development of different meanings influenced the retention of the two stems needs to be investigated in more detail, however.

This difference was explained through the process of reanalysis, whereby Sakha forms that from a hearer's perspective have an ambiguous morphological structure, were divided up in a different way by speakers of Dolgan. While this cannot be determined with absolute certainty, for relational nouns it is most likely

⁵ Original transcription: *aatta* (Sauer 1803: 318).

⁶ Original transcription: *tina*, translated as 'ghost, soul' (lit.: Geist, Seele) (Sauer 1803: 320).

that the high final vowel in the Dolgan stem corresponds to the high vowel that appears in the possessive case suffix of the third person singular. Since relational nouns as a default figure in the *izafet* construction of locational phrases, and often have third person referents, this form occurs most frequently in discourse and is therefore most likely to have served as a basis for this reanalysis. For the referential nous, the most plausible origin of the final high vowel was argued to be the epenthetic vowel that is inserted between stem and suffix in all possessive forms but the third person singular⁷.

5.3. REGULARISATION OF THE PARADIGM OF AUXILIARY VERB *e*- ‘BE’

5.3.1 DESCRIPTION

Another instance of paradigm regularisation in Dolgan is the declension of the defective auxiliary verb *e*- ‘be’. Like in Sakha, this verb is only used in the past tense in Dolgan, and is employed as: a) an auxiliary verb to form analytical past tense forms (resultative past, the imperfective past and the remote past); and b) as a copula with nominal predicates. Both uses are illustrated in example 5.20.

- (5.20) *hoɣoto:χ* *kii:s* ***e-ti-m*** *buo* *kergetter-ber*,
 single girl be-PST-POSS.1SG PRT family.PL-DAT.1SG
χahan *da* *giniler-ten* *araχ-pataχ* ***e-ti-m***
 when NEG 3PL-ABL leave-PST.PTC.NEG be-PST-POSS.1SG
 ‘Well, I was a single child for my parents, I had never been separated from
 them’ (LKS: 24)

In normal verbs such as *bar* ‘go’, most person-number forms of the recent past are formed according to the scheme: STEM+*TI*+POSS, whereby *TI* is the recent past suffix, and POSS the possessive person-marking suffix which agrees with the subject. However, the third person deviates from this scheme in the singular and plural. A regular formation of *TI*+POSS for the verb *bar* ‘go’ would look like **bardita* [*bar-TI-(t)A*, go-PST-POSS.3SG.] for the singular, and **bardilara* [*bar-TI-LArA*, go-PST-POSS.3PL] for the plural. However, Table 5.13 shows that the attested

⁷ As was shown in Table 5.12, only in the nominative case this leads to a low final vowel. In all other cases the vowel following the assimilated stem is also high, but from a morphological point of view this high vowel belongs to the suffix, and is no epenthetic vowel, as is the case in other persons.

forms are *barda* and *bardilar*. For the third person singular it could still be argued that it follows the regular scheme if the past tense suffix is analysed as *-T(I)* instead of *-TI*, that is when the high final vowel is seen as an epenthetic vowel, which only appears when the following suffix begins with a consonant, as in *bardibit* [*bar-T(I)-Bit*, go-PST-1PL]. In this case, the analysis of *barda* would be *bar-T(I)-(t)A*, go-PST-POSS.3SG, and would be regular despite the absence of a high vowel in the past tense suffix.

However, the third person plural form is not so easy to explain within the regular paradigm. Even an analogical division of the suffix *-TilAr* into past tense suffix *-T(I)* and plural suffix *-LAr* does not make it compatible with the regular scheme of stem+PST+POSS, according to which we would expect a third person plural form of *bardilara* [*bar-T(I)-LArA*, go-PST-POSS.3SG]. An example of the entire glossed paradigm for the recent past is presented in Table 5.13, alongside with the inflectional paradigm of a possessive marked noun for comparison. The difference in person marking is highlighted in bold.

Table 5.13: Inflectional paradigm for recent past and nominative possessive declension

RECENT PAST OF BAR 'GO'		POSSESSIVE DECLENSION OF <i>aya</i> 'FATHER'	
<i>bar-d̥i-m</i>	'I went'	<i>aya -m</i>	'my father'
<i>bar-T(I)-(I)m</i>		<i>aya -(I)m</i>	
go -PST -POSS.1SG		father -POSS.1SG	
<i>bar-d̥i-ŋ</i>	'you went'	<i>aya-ŋ</i>	'your father'
<i>bar -T(I)-(I)ŋ</i>		<i>aya -(I)ŋ</i>	
go -PST -POSS.2SG		father-POSS.2SG	
<i>bar-d-a</i>	'he went'	<i>aya-ta</i>	'his father'
<i>bar -T(I)-(t)A</i>		<i>aya -(t)A</i>	
go -PST.3SG		father POSS.3SG	
<i>bar-d̥i-bit</i>	'we went'	<i>aya-bit</i>	'our father'
<i>bar -T(I)-Bit</i>		<i>aya -Bit</i>	
go -PST -1PL		father-1PL	
<i>bar-d̥i-git</i>	'you went'	<i>aya-git</i>	'your father'
<i>bar -T(I)-Git</i>		<i>aya -Git</i>	
go -PST -2PL		father-2PL	
<i>bar-d̥i-lar</i>	'they went'	<i>aya-lara</i>	'their father'
<i>bar -T(I)-LAr</i>		<i>aya-LArA</i>	
go -PST -3PL		father-POSS.3PL	

In Sakha, the past tense of the auxiliary verb *e-* ‘to be’ is inflected in exactly the same way as *bar* or any other verb, with the third person singular and plural following the ‘regular’ irregular pattern:

SAKHA

- (5.21) *onon bu xaray-ım uruk-ka-ttan möltöx e-te*
 therefore this eye-POSS.1SG in.past-ADJZR-ABL weak be-PST.3SG
 ‘And this eye was weak even before.’ (ESY: 98)

- (5.22) *Hür-de:χ üčügej, akti:binaj oyo-lor e-ti-ler [...]*
 very.much -PROP good active child -PL be -PST -PL [...]
 ‘They were very good, active kids [...].’ (AGM: 177)

In Dolgan, however, the form of the third person plural is changing. Instead of using the typical suffix *-T(I)-LAr*, the overwhelming majority of third person plural forms in the spoken corpus appear as *etilere* [*e-T(I)-LArA*, be-PST-POSS.3PL] ‘they are’.

DOLGAN

- (5.23) *min ha:-la:χ e-ti-m beje-m, onton*
 1SG gun-PROP be-PST-POSS.1SG self-POSS.1SG then
doyottor-um it-tar-da:χ e-ti-lere
 friend.PL-POSS.1SG dog-PL-PROP be-PST-POSS.3PL
 ‘I had a gun myself, and my friends had dogs’ (SEK: 10)

Out of 55 occurrences of third person plural forms in the corpus, 47 are *etilere* (85.5%), and only 8 (14.5%) correspond to the Sakha form *etiler*. In the Sakha corpus, which consists of several Sakha dialects from a wide range of geographical regions including the Olenek region, which is relatively close to the Taimyr, not a single instance of *etilere* was recorded. In contrast to the third person plural, the third person singular is identical in both languages, which would support an analysis of *ete* as a regular form and thus for the past tense suffix to be of the form *-T(I)*. An overview of the paradigm for *e-* ‘be’ for both Dolgan and Sakha is given in Table 5.14.

Table 5.14: Recent past for auxiliary verb *e-* ‘be’

SAKHA	DOLGAN	
<i>e-ti-m</i>	<i>e-ti-m</i>	‘I was’
<i>e - T(I) -(I)m</i>	<i>e - T(I) -(I)m</i>	
<i>e-ti-ŋ</i>	<i>e-ti-ŋ</i>	‘you were’
<i>e - T(I) -(I)ŋ</i>	<i>e - T(I) -(I)ŋ</i>	
<i>e-t-e</i>	<i>e-t-e</i>	‘he was’
<i>e - T(I) -(t)A</i>	<i>e - T(I) -(t)A</i>	
<i>e-ti-bit</i>	<i>e-ti-bit</i>	‘we were’
<i>e - T(I) -BlIt</i>	<i>e - T(I) -BlIt</i>	
<i>e-ti-gīt</i>	<i>e-ti-gīt</i>	‘you were’
<i>e - T(I) -GlIt</i>	<i>e - T(I) -GlIt</i>	
<i>e-ti-ler</i>	<i>e-ti-lere</i>	‘they were’
<i>e - T(I) -LAr</i>	<i>e - T(I) -LArA</i>	

5.3.2. DIRECTION OF CHANGE AND EARLIER ACCOUNTS IN THE LITERATURE

With respect to the direction of change, we can be confident in assuming that Dolgan is the innovative language. Comparative data from other Turkic languages show that a past tense category with a suffix related to *-T(I)-LAr* is very common within the Turkic language family (including Dolgan for all verbs except *e-* ‘be’). To substantiate this, Old Turkic uses *-dīlAr* in the constative preterite (Erdal 1998: 145, 2004: 327), Turkish and Azerbaijani use *-DI-lAr* for the simple past (Csató and Johanson 1998: 214, Schönig 1998: 254), and Tatar and Bashkir *-DĖ-lAr* (Berta 1998: 292).

As in the case of the unstable stems discussed above, grammatical descriptions of Dolgan provide contradictory information with respect to the inflectional paradigm of *e-*. According to Ubryatova (1985: 167), the paradigm is identical to the one in Sakha, i.e. the third person plural is *etiler*. As far as I am aware, she makes no comments on possible variation of this form in written or spoken discourse. Artemyev (2001: 196), on the other hand, gives only the Dolgan form *etilere*. However, in some of his examples elsewhere in the grammar the form *etiler* also appears (ibid.: 200). Despite the presence of these two forms in Artemyev’s data, I have not been able to find a discussion of this variation.

There are several possible reasons for the divergence in these descriptions. First, to do both authors justice, it may be the case that the innovation in Dolgan is

relatively recent. Ubryatova's grammar is based on linguistic material collected in the 1930s, which leaves a significant time span of about 60 years before Artemyev published his grammar in 2001. The fact that Sakha *etiler* is still in use as well, despite the current dominance of *etilere*, could be interpreted as supportive evidence for the idea that this is a recent innovation in Dolgan, that is, both allomorphs are still used and the new form has not yet taken over.

A second explanation may lie in the fact that Ubryatova is a specialist on Sakha as well. Having a thorough knowledge of a very closely related language has many advantages, but could potentially lead to false assumptions about language Y (the new language) on the basis of language X (the language already known). In the case of *etiler* this is not unthinkable. First, only this defective auxiliary verb shows the allomorphy of *-T(I)-LAr* and *-T(I)-LArA*, therefore this variation would not show up while studying verbal morphology with other verbs. Second, it is certainly not the case that *etiler* in Dolgan is ungrammatical. Upon explicit asking whether there is a difference in use or meaning between the two forms, speakers refute that option and say they are fully equivalent and interchangeable. Only a corpus count shows that in practice *etilere* is evidently favoured in spoken discourse. Thus, depending on the way Ubryatova's language material was collected, she may not have come across the form, even if it were already present in Dolgan, because Dolgan speakers would never have rejected *etiler* as a form not belonging to their verbal paradigm.

5.4. DISCUSSION

This chapter has presented two instances of regularisation of paradigms. The first case was classified as an example of reanalysis, the second as regularisation on the basis of analogy with the paradigm of possessive person marking. In the beginning of this chapter, reanalysis was introduced as a process by which the underlying structure of a morphosyntactic sequence changes over time, while the surface structure remains largely unchanged. It was stated that this occurs as a result of potential structural ambiguity of a certain surface form through mechanisms of analogy. The data in this chapter have shown that this is what has happened in unstable stems in Dolgan and that reanalysis is an appropriate term to describe the differences between Dolgan and Sakha in this domain. But how common is this kind of change in languages, and how can it be explained? The next section gives

an overview of what has been said about this topic in terms of language-internal as well as language-external motivations for reanalysis.

5.4.1 PRINCIPLES UNDERLYING PARADIGM REGULARISATION AND REANALYSIS

Within the literature on language development and language change, reanalysis occupies a prominent place. This holds for syntactic as well as morphological change. According to Harris and Campbell (1995: 3) reanalysis has been “(...) the single most important mechanism for most attempts to explain syntactic change throughout the history of linguistics.” Similarly, Joseph (1998: 358) describes the emergence of morphology as follows:

The primary source of morphology is material that is already present in the language, through the mediation of processes of resegmentation and reinterpretation applied in a variety of ways, as well as by other processes of change – for example sound changes – that lead to grammaticalization.

An appealing account of the cognitive principles and mechanisms that may underlie the process of reanalysis is provided by Bybee (1991). In her word-based model of morphological organisation, with a focus on the organisation of paradigms, she stresses the crucial role of language use in shaping language structure. She postulates a strong link between the way language forms appear in discourse and the mental representations that underlie them.

While her account in the 1991 paper deals mainly with the acquisition of morphological paradigms, Bybee’s theory suggests a direct link between the acquisition process and the restructuring of morphological paradigms, sometimes through reanalysis. Thus, she assumes that language learners (of L1 as well as L2) are directly involved in language change because they hold assumptions about paradigm structures which may be different from their predecessors or peers.

For Bybee, essential concepts in the dynamics of language are ‘basic form’, ‘markedness’ and ‘frequency of use’. These three concepts are closely intertwined in the language learning process and in the determination of directionality in paradigm restructuring. ‘Basicness’ of a form is determined by two factors: a) high frequency in discourse; and b) semantic unmarkedness. Typically semantic unmarkedness correlates with morphological unmarkedness and the idea is that

the marked forms in the paradigm are derived from the basic unmarked form. Typical examples of basic grammatical categories are singular, nominative, first and third person, present tense and indicative mood.

However the correlation between semantic and morphological markedness does not always hold, and can be reversed depending on the frequency of use of the marked form. Sometimes semantically and morphologically marked forms can become 'basic' and thus serve as a basis for regularisation of a morphological paradigm. This phenomenon is called 'local markedness', which I will focus on in some detail because it is a relevant concept in the explanation for the paradigm differences between Dolgan and Sakha. Tiersma (1982: 833-834) provides examples of Frisian, where the plural stem of certain words has been generalised over the entire paradigm to form the basis of the singular too. By the same token, Mańczak (1980: 285) illustrates local markedness with the dominance of locative case in place names and the instrumental case of nouns designating tools. While 'local markedness' seems to go against the idea that basic forms correlate with the semantically and morphologically unmarked members of the paradigm (after all, in the examples referred to, it was nouns marked for plural, locative and instrumental that came to serve as the basis of the morphological paradigm), in all these cases the marked form has a much higher token frequency than the unmarked forms. The plural stems in Frisian were all nouns that normally occur as pairs or in groups in natural discourse, such as *arm*, *tooth*, or *tear*. Similarly, for obvious semantic reasons place names occur most frequently in the locative case and nouns denoting tools in the instrumental case. Therefore the 'marked' categories are for these words more 'basic' than the unmarked.

Bybee concludes that whatever is inherent in the meaning of a word (such as the plural for *tooth*, or the instrumental for *knife*) is treated cognitively as unitary, non-complex and 'basic'. What is marked and unmarked may therefore depend on the semantic properties of an individual lexical item. That is, the plural in *noses* is marked since an individual normally has only one of them, but the plural in *lice* is typically unmarked since they unfortunately tend to appear in large quantities. However, in all cases there is a strong correlation between semantic unmarkedness (singular for *nose* and plural for *louse*) and frequency of use. Thus, frequency of use has the potential to reverse the correlation between semantic and morphological unmarkedness. It strongly influences what is conceived of as 'marked' and 'basic' and is therefore a major and overruling factor in patterns of paradigm levelling.

With respect to language acquisition and learning strategies, a higher token frequency means a higher proportion in the input and therefore a stronger representation in the mental lexicon. Therefore the most frequent forms are learned first, and during language acquisition the rest of the paradigm is interpreted as being derived from these forms. Normally this is the morphologically and semantically unmarked basic form, but in the case of ‘local markedness’ a morphologically marked form can serve this purpose equally well. Although Bybee does not provide support for this claim with experimental data, she assumes on the basis of the general cognitive nature of these learning principles that the same learning strategies and generalisations take place in second language learning.

5.4.2 REGULARISATION AND REANALYSIS IN DOLGAN EXPLAINED

The previous section has provided a definition of regularisation in morphological paradigms by reanalysis, and an understanding of the cognitive mechanisms that may underlie it, with a focus on language-internal development and L1 acquisition. In this section I will discuss how paradigm regularisation in Dolgan can be interpreted in terms of these mechanisms.

With respect to these changes there are two questions to be answered: a) how did this change happen in Dolgan; and b) why did it not happen in Sakha? After all, if reanalysis and regularisation are such common language-internal processes, there is no obvious reason why the same forms would not have developed in other Sakha-speaking regions. In this section I will deal with question a); question b) will be discussed in the next section.

For relational as well as referential nouns it was shown that the ambiguity of forms has led to a change where the oblique stem has become generalised over the entire paradigm, with a few exceptions of very frequent, possibly fossilised forms, where the basic Sakha stem is still used. As Koch explains it, in ambiguous situations the underlying form that will eventually be selected is ‘the word form that appears most frequently for the particular lexeme’ and has the highest ‘paradigm frequency’ (Koch 1996: 232), that is the number of slots in the paradigm in which it occurs. Typically, these conditions correlate with Bybee’s prototypical basic form, in which high frequency of use, and semantic, and morphological unmarkedness conspire towards an ideal basic form. However, as was illustrated

with the example from Frisian, ‘local markedness’ provides an exception to this correlation, and an important cue to how the semantics of individual lexical items can shift the centre of gravity in the understanding of ‘basicness’. The regularisation in Dolgan seems to be an illustration of exactly this.

The relational nouns illustrate this in an obvious way. Like postpositions, their inherent (read: unmarked) semantics represent locational, directional or instrumental relations, which in Sakha correlate with possessive marked dative, ablative and instrumental cases. Consequently, these are the forms of relational nouns that are expected to occur most frequently in discourse, which was unequivocally confirmed by the data from the Sakha and Dolgan corpus, in which only very few unmarked nominatives were found. Whether it was high frequency that triggered this interpretation of semantic unmarkedness or the other way round remains a chicken and egg type question, but it is clear that for relational nouns there is a strong correlation between the two. The high frequency and semantic unmarkedness qualify the morphologically marked forms to be interpreted as basic.

For the referential nouns the picture is very similar. Again, the oblique stem has become generalised over the whole paradigm, and judging by the high final vowel in Dolgan stems today, in this case forms taking an epenthetic vowel lie at the heart of this reanalysis, which is all possessive marked forms, except the third person singular. This nicely fits the idea of ‘local markedness’ discussed in 5.4.1. In many cases the reanalysed referential nouns concerned concepts (family members, body parts) that for semantic reasons typically require possessive marking. Therefore, the possessive-marked form is for these nouns the semantically unmarked, the most frequent and therefore most basic form. This makes it another illustration of ‘local markedness’ and explains how a morphologically marked form can become the basic form within a paradigm. More generally, it illustrates the importance of semantic properties in the frequency of occurrence of morphological features, and thus in the structuring of morphological paradigms.

Furthermore, it was demonstrated that some forms continue to exist in their Sakha shape, even though the oblique form in Dolgan clearly dominates the morphological paradigms in the category of unstable stems. For relational nouns, it was shown that this could be explained in terms of discourse frequency as well. As a reminder, *kenne*, *kennitten*, *ürdüger* and *ürdünen*, which are the Sakha forms that occur in Dolgan discourse, are the most frequently used forms in Sakha. For

these particular forms I argue that there are forms such as *kennitten* that, due to the high input frequency, have become stored in the mental lexicon as an unanalysable unit, and continue to exist alongside the reanalysed forms, where the relational nouns are taken apart into a stem and a case suffix. In the case of Dolgan, they seem to occur in free variation.

For referential nouns, the preservation of the Sakha forms may be explained along similar lines. However, an additional factor which may have increased the ambiguity in the first place is the apparent randomness of the application of the phonological rule that eliminates the final vowel in this word class (i.e. the difference between stable and unstable stems). The fact that there are some words for which this rule applies and others for which it does not, may have enhanced uncertainty with respect to the underlying form, which could have lead to the use of both stems, and the preservation of Sakha stems in some nouns. Over time the ‘one form one meaning’ principle could have lead to a differentiation in meaning between the two forms, as we see for *tumus*.

5.4.3 THE POTENTIAL ROLE OF L1 AND L2 LEARNERS IN REANALYSIS

The previous sections have shown that reanalysis and regularisation of paradigms is a common phenomenon in natural language change and can often be satisfactorily explained in terms of language-internal motivations. In other words, from a purely linguistic point of view there is no obvious need to include the influence of second language learners or language contact in the picture. Admittedly, distinguishing between L1 and L2 influence is a difficult task, since principles such as markedness and frequency are of a general cognitive nature, and presumably apply to learning mechanisms in L1 as well as L2 acquisition (Bybee 1991: 88). However, an exclusively language-internal account of the changes in Dolgan morphology leaves one issue unexplained, namely their geographical distribution. If paradigm regularisation is such a common language-internal development, why did this particular change happen only in Dolgan and not in any dialects of Sakha? All things being equal, one could expect a linguistic change to arise and gradually spread across the speech community if the social conditions are favourable. However, the changes described here are restricted to the Dolgan-speaking area only.

Since the cognitive conditions are obviously the same for speakers of Sakha on the Taimyr and, say, speakers of Sakha in Central Yakutia, more significance must be attributed to the equally important set of explanatory factors in language change, namely the sociolinguistic situation. It is widely accepted that the social situation in which a change occurs is crucial for its further spread and development within the speech community (see Section 3.1.4 for discussion). We know that the sociolinguistic situation was certainly not identical across Sakha-speaking communities, so therefore it may have conditioned the appearance and spread of certain linguistic variants within particular parts of the broader speech community.

An important sociolinguistic condition to consider within the context of this dissertation is the potential interaction with other populations. It appears that there is a correspondence between the area where paradigm regularisation has taken place (the Taimyr Peninsula) and where we know from historical records that different populations have been in contact (in particular Sakha and Evenks). Based on historical sources as well as on research in other linguistic domains, such as the lexicon (see Chapter 4) we may assume interaction between these populations and a certain degree of bilingualism. Although historical evidence of population contact alone is no guarantee for significant linguistic contact, the overlap between the area of population contact on the Taimyr and the spread of this particular change increases the potential relevance of second language learners in the rise of this change in Dolgan. Recognising this role, one could imagine that paradigm regularisation came about through L1 Evenki speakers, who were learning Sakha as a second language. During the learning process they may have reanalysed and regularised the paradigms of ambiguous noun stems, motivated by cognitive principles related to semantic and morphological markedness and frequency, as discussed above.

Thomason and Kaufman (1988: 145) describe paradigm regularisation, or simplification, as a phenomenon that is characteristic for L2 acquisition as much as it is for L1 acquisition, and associate it with a situation of language shift. Where in monolingual Sakha communities regularised variants in young children in all likelihood get corrected over time (presuming that they do occur in L1 acquisition as well), this has not happened in the Dolgan community. This difference may have several explanations. First, it may be due to the age of the bilingual speakers. It is commonly known that the ability to achieve native-like fluency in a second language decreases with age, and becomes impossible after the so-called critical

period (see Chapter 3, Section 3.1.3). While leaving the details as to the exact age limit unaddressed, the important point is that non-native variants are much more likely to remain uncorrected in adult bilinguals than in children. Since it is hard to imagine a shifting population that consisted only of children, it is plausible to assume a large number of bilingual adult speakers, leading to a stronger persistence of regularised variants.

Second, the regularised forms may have become established due to a large number of shifting speakers. When the group of shifting speakers is large, non-native variants are more likely to stay within the community than if only a few speakers were shifting. In a large group the bilingual speakers are more likely to hear each others' foreign version of the language they are shifting to than when the shifting group is small, and consequently their exposure to 'native' variants of Sakha would be lower than for small shifting groups.

A third important point is the fact that in the contact setting on the Taimyr Sakha was used as a lingua franca. This means that even without taking into account the shifting groups, there were many second language speakers of Sakha. Since regularisation is known to occur frequently in languages of wider communication, the function of Sakha on the Taimyr in this capacity is another plausible explanation for the geographical distribution of the change.

A final possible explanation may be found in current L1 attrition. Attrition often involves simplification due to the lack of language use, language input, or the lack of exposure to the language, which causes details and irregularities to be levelled out. Attrition can affect every linguistic subdomain, including the structure of a speakers' L1. As Sharwood Smith and Van Buren (1991: 20) have it,

...the attrition of competence may be triggered by changes in the learner's perception of the basic structure of his or her L1 grammar, and not by a tendency to ease the processing burden of an underused L1."

It is clear that it is impossible to categorically tease apart the influence from L1 and L2 acquisition because many of the cognitive principles apply to both. Thus it will never be possible to prove whether regularisation in Dolgan paradigms is the result of language-internal change of language shift, or of the use of Sakha as a lingua franca. However, if contact played a significant role in the development of these differences between Dolgan and Sakha, it would most likely be caused by considerably large groups of adult non-Sakha speakers, who shifted to Dolgan/Sakha and learned it as a second language. However, language-internal

development is probable too, since regularisation is common in internally motivated language change.

As unrealistic as it is to have the desire to completely separate language-internal and language-external motivations, so it is unrealistic that these factors were radically separated in reality. Rather I would argue that the linguistic outcome we see in present-day Dolgan is the result of the interplay between the two scenarios, which reinforced each other. In L1 and L2 acquisition, regularisation occurs as a consequence of general cognitive learning principles in a situation of plausible language change. If a regularisation was made by a second language learner, it may have become more easily accepted by native speakers because it is a plausible change for L1 speakers as well. Similarly, L1 speakers who regularise paradigms during the acquisition process may hear these forms from other people around them and thus the language-internal tendency would be reinforced. Thus it is important to take both factors into account as significant possibilities in the explanation of the divergence of Dolgan and Sakha in this respect, even if they are technically inseparable.

6.1 INTRODUCTION

In this chapter the habitual aspect in Dolgan is investigated. On the basis of spoken text corpora for Dolgan and Sakha, I will show that in Dolgan a) the frequency of use of the habitual participle is significantly higher than in Sakha; b) the habitual participle is used predominantly with a verbal function; and c) the nominal function is, counter to grammatical descriptions, virtually absent. The second part of the chapter focuses on the cause of these differences, and the possibility of both language-internal and language-external explanations will be considered. While no definite conclusions can be reached at this stage due to gaps in the data, hypotheses are formulated that uncover important areas for future research.

In this cross-linguistic analysis of habitual aspect, I have used semantic rather than morphosyntactic criteria for what is considered habitual aspect. According to Comrie

[a]spect is not concerned with relating the time of the situation to any other time-point, but rather with the internal temporal constituency of the situation.
(Comrie 1976: 5)

This sets it apart from tense, which is “grammaticalised location in time” (Comrie [1985] 2000: 9). Habitual aspect in particular is defined as:

A situation which is characteristic of an extended period of time, so extended in fact that the situation referred to is viewed not as an incidental property of the moment but, precisely, as a characteristic feature of the whole period. (Comrie 1976: 27)

Although tense and aspect are independent categories, it appears from cross-linguistic study that overt marking of habitual aspect is associated much more with the past tense than with the present (Bybee, Perkins & Pagliuca 1994). This tendency is explained by the idea that habitual aspect “may be one of the basic or default aspectual readings of the present tense” (ibid.: 151), and may therefore often not be overtly expressed in such contexts.

Habitual aspect should not be confused with iterative aspect, which is the “successive occurrence of several instances of the given situation” (Comrie 1976: 27). As Comrie points out, iterativity does not imply habituality or the other way round. The repetition of an event does not necessarily give it a habitual character, as in *the lecturer stood up, coughed five times, and said...* (Comrie 1976: 27) and habitual events do not necessarily involve iterativity as in *Simon used to believe in ghosts* (ibid.). However, habituality and iterativity may be combined in one event as in, for example, *before he started his lecture, the lecturer used to cough five times*. In this sentence, the *five times* encodes the iterativity of the event of coughing, whereas *used to* indicates its habituality. However, as will be clear from the preceding examples, the two categories are independent of each other. In the remainder of this chapter I will consider habitual aspect only.

6.2 HABITUAL IN DOLGAN AND SAKHA AND THEIR POSITION AMONG OTHER TURKIC LANGUAGES

Within the Turkic language family, the habitual aspect is expressed in a variety of ways. In Dolgan and Sakha, it is formed with the suffix *-A:ččI* followed by predicative person marking that agrees with the subject¹.

¹ In Sakha, *-A:ččI* also occurs in the function of agent nominaliser (e.g. *kömölöh-öiččü* [*kömölös--A:ččI*, *help-HAB*] ‘helper’). See section 6.3.2 for discussion.

DOLGAN

- (6.1) *ol kördük (...) emte-n-e:čči-ler ke*
 DEM similar (...) cure-RFL-HAB-PRED.3PL PRT
 ‘This is how they tend to cure.’ (IMA: 44)

SAKHA

- (6.2) *dʒoŋ-ŋo kömölöh-ö:ččü-bün [...] daja:rka-lar-ga inax*
 people-DAT help-HAB-PRED.1SG [...] milker-PL-DAT cow
torbos ejin kömölöh-ö:ččü-bün tard-ih-a:čči-bün,
 calf etc. help-HAB-PRED.1SG pull-RECP-HAB-PRED.1SG
ah-a-t-ih-a:čči-bün.
 food-VBLZR-CAUS-RECP-HAB-PRED.1SG
 ‘I help people, I help the milkers (with the) cows and calves and so on, I
 tend to help pull, I help feed.’ (adapted from XLE: 188)

In the past tense, the habitual participle is unmarked for person, and subject agreement is expressed on the auxiliary verb *e-* ‘to be’ by means of a possessive suffix.

DOLGAN

- (6.3) *hohuj-a-bün, küččügüj kihi ke,*
 be.frightened-SIM.CV-PRED.1SG small person PRT
kuttan-a:čči e-ti-m buo
 be.scared-HAB be-PST-POSS.1SG PRT
 ‘I was frightened, I was small you see, I was always afraid.’ (TJP: 14)

SAKHA

- (6.4) *Üčügej bayajı buol-a:čči e-t-e*
 good very AUX-HAB be-PST-POSS.3SG
 ‘It was very good.’ (XKM: 095)

This encoding strategy is unique within the language family. In other Turkic languages, the habitual function is expressed by means of different suffixes. For example, in Old Uyghur and Qarakhanid the habitual participle in *-gAn* is used to express this meaning². According to Erdal, the suffix *-gAn* was obsolescent in Old

² In other Turkic languages such as Khakas and Kyrgyz this suffix is used to form the past tense.

Turkic³, but became more and more productive over time. In many modern Turkic languages it nowadays occurs as a participle and in the formation of action nouns (Erdal 2004: 156, 290, Erdal 1991: 387).

In Khakas the present habitual is formed with the suffix *-AdVr/-idVr*, as can be seen in example 6.5, and which has its origins in the converbal forms ending in *-A* or *-i* respectively, followed by the auxiliary *tur-*. The past is formed with the suffix *-žarj*.

KHAKAS

- (6.5) *kem-nejer čooxt-an-za-ŋ ol kIr-edIr*
 who-CIR speak-RFL-COND-2 he enter-HAB.PRS
 ‘Whoever you might be talking about he always shows up.’
 (adapted from Anderson 1998: 40)

In Turkish the habitual function is fulfilled by the aorist suffix *-(V)r*. Kornfilt describes it as the “general present tense [which] expresses habitual actions and general events, thus coming close to a universal tense” (Kornfilt 1997: 336).

TURKISH

- (6.6) *Hasan her sabah kahvaltı ed-er*
 Hasan every morning breakfast do-AOR
 ‘Hasan has breakfast every morning.’ (Kornfilt 1997: 336)

In Kyrgyz and its close relative Altay⁴ the suffix *-U:čU* (mostly occurring as *-čU*) is found (Somfai Kara 2003: 32, 43, Kałużyński 1995). These suffixes function as agent nominalisers, as well as habitual participles in Kyrgyz (Somfai Kara 2003: 43).

³ However, Erdal mentions that even in late Old Turkic there is evidence that this verb form was used as a participle (Erdal 2004: 156).

⁴ According to Johanson’s classification of Turkic languages, the classification of Kyrgyz is ambiguous. It may be classified as belonging to the southern subbranch of the northeastern group, to which Altay also belongs, but recent changes have made it more similar to Kazakh, which is part of the southern subbranch of the northwestern group. Therefore, some scholars classify it within that group (Johanson 1998: 83).

KYRGYZ

(6.7) **kel-(ü:)-čü (ele)**
 come-HAB be-PASS

‘He used to come.’

(Somfai Kara 2003: 43)

The functional and formal similarity with *-A:ččI* makes relatedness of these two suffixes a tempting hypothesis. However, Kałużyński (1995), after Ramstedt, argues against this, saying that *-U:čU* can be traced back to a combination of the Turkic suffixes *-yg+čy*, in which case the sound correspondences with Sakha *A:* do not fit. According to Kałużyński, the sound combination *-yg* normally corresponds to *-i:* or *-ia* in Sakha, and not to *-A:*.

As can be seen from this brief overview, Dolgan and Sakha are exceptional within the Turkic language family in using the participle *-A:ččI* to express habitual aspect. The next section will show that related forms of this suffix are found in a few other languages of the northeastern branch of the Turkic family. However, their use remains restricted to the nominal realm, in particular to the function of agent nominaliser, and in none of them has it acquired the function of a habitual aspect.

6.3 THE HABITUAL IN *-A:ččI*

6.3.1 THE ORIGIN OF *-A:ččI*

According to Korkina (1970: 220) and Kałużyński (1995: 101), citing Ramstedt (1952), the suffix *-A:ččI* is found only in Tuvan, Khakas, Altaic and Sakha; that is, only in the north-eastern branch of the Turkic language family. In all these languages the suffix is used as an agent nominaliser, as in Khakas *ojnāči* ‘player’ from the verb stem *ojnā* ‘to play’ (Kałużyński 1995: 101).

Speculations about the origin of the *-A:ččI* suffix are rather divergent and not always equally convincing. For example, in Korkina’s overview on the origin of the suffix she cites Böhtlingk’s (1851: §722) suggestion that *-A:ččI* may have come from the agent noun in *-I:*, whereas Khitrov (1858: 121) proposes a relation to the Sakha converb in *-n*, which would have subsequently been exchanged for the suffix *-ččI*. His motivations for this unusual replacement, or the origins of *-ččI* itself, are not further specified. Finally Korkina (1970: 225) refers to Radloff (1908: 50) who, conversely, relates *-A:ččI* to the converb in *-A*. Korkina herself puts forward that

the meaning of generality and permanency attributed to the habitual participle in Sakha can be considered a more intense variant of the meaning conveyed by the present participle in *-Ar*, which motivates her hypothesis that *-A:ččI* may have its origin in a combination of the present participle suffix *-Ar* combined with the nominaliser suffix *-čI*. Its allomorph *-A:ččIk* she traces back to the Russian agent nominaliser suffix *-sh'ik*. However, she leaves the reason for the variation, as well as for the connection with the Russian suffix, unexplained. Ubryatova (1985: 184) objects to this analysis in her description of the habitual participle for Dolgan, arguing that the relation with the present participle is not proven. She prefers to reconstruct the form as a combination of the Turkic suffixes **-gač* and **-či*, for reasons that are not entirely clear. Remarkably, she suggests a different origin for *-A:ččIk*, even though she acknowledges it as just a phonetic variant of *-A:ččI*. This in itself being rather unusual reasoning, she reconstructs the origin of *-A:ččIk* as a combination of **-gač* and **-erjik* without further clarification.

These language-internal explanations are rather opaque due to their divergent character and an often inapparent correspondence between the current form and its hypothesised components. In contrast, Kałużyński (1995), after Ramstedt, offers a language-external, and more plausible, explanation. It appears that the nominal use of a suffix with a very similar form, *-yači*, is also found in Mongolic languages, including Kalmykian, Mongol proper (including Khalkha), and Written Mongol, which leads Kałużyński, again following Ramstedt's argumentation, to the conclusion that the suffix must have been copied from Mongolic into the Turkic language family. They argue that *-A:ččI* has its origins in Proto-Mongolic **xA.ci*, which was the marker of the agentive participle (Janhunen 2003: 21). Poppe mentions a related suffix in his grammar of Written Mongolian, where he describes the function of *-yači* as "...to form nouns designating names of vocations" (Poppe 1991). According to this scenario, this function, along with the form, has been copied into the Turkic languages Tuvan, Khakas and Altaic and Sakha, in which the suffix is found in that function today. This explanation seems probable, considering the fact that many of the Turkic languages in which the suffix is found are spoken in the area bordering present-day Mongolia and it is known from history that this area was dominated by Mongolic-speaking people for a long time.⁵ Since in this account the correspondence in form as well as in

⁵ Although it is hard to give exact dates, contact between speakers of Mongolic and Turkic languages goes back to the second half of the first millennium AD. It intensified in the 11th and 12th centuries when certain Mongolic tribes fled north to avoid internal conflicts and it is assumed that they arrived

meaning is more transparent than for the available language-internal explanations, the idea that *-A:ččI* has its origin in Mongolic languages is more likely.

6.3.2 USE OF *-A:ččI* IN DOLGAN AND SAKHA

In contrast to the other north-east Turkic languages, in which the forms related to *-A:ččI* have the function of an agent nominaliser, in Sakha and Dolgan its use is predominantly verbal. This tendency is particularly pronounced in spoken Dolgan, where the nominal use seems to be completely absent.

Despite the overwhelming percentage of verbal use that emerged from a frequency count of the spoken Sakha corpus (see section 6.3.3.4) most of the literature on Sakha focuses on the nominal and adjectival use of the participle in *-A:ččI*. Böhlingk ([1851] 1997: §722) writes that “[t]he verbal noun in *-A:ččI* only ever occurs as an agent noun and is used adjectivally in combination with a noun, and as a noun”⁶. A similar view is held by Kharitonov (1947), as well as by Khitrov (1858: 121) and Poppe (1926: 67) as cited by Korkina (1970: 220), who all highlight the nominal aspect of the participle, reflected in descriptions such as ‘verbal noun’ or ‘agent noun’. Indeed, this is how the participle can be used in contemporary Sakha, as can be seen in example 6.8. The adjectival use is illustrated in example 6.9.

at the area around Lake Baykal, where they may have met the Turkic-speaking Sakha. A subsequent peak was during the Mongol Empire in the 13th and 14th centuries, when more Mongolic clans are supposed to have entered the area of Lake Baykal to escape from the power of Chinggis Khan (Pakendorf 2007: 22-23).

⁶ “Das Verbalnomen auf *-ачы* ist immer nur Nomen agentis und wird sowohl adjectivisch in Verbindung mit einem Substantivum, als auch substantivisch gebraucht.” O. Böhlingk (1851: § 722 (English translation mine)).

SAKHA

- (6.8) *dze mama-bar ha:maj tireχ buol-an χa:l-l-īm,*
 Well mother-DAT.1SG the.most support AUX-SQ.CV remain-PST-POSS.1SG
kömölöh-ö:ččü, (...) buol-an χa:l-l-īm.
 help-HAB (...) AUX-SQ.CV remain-PST-POSS.1SG
 ‘Well I remained my mother’s biggest support, I remained her helper.’
 (ARR: 031)

- (6.9) *Ol tojon halaj-a:ččī kihieχ aya-bar*
 that chief lead-HAB person.DAT father-DAT.1SG
ki:l-ler-el-ler bu kihi-ler-in bandži:t-tar.
 enter-CAUS-PRS.PTC-PL this person-PL-ACC.3SG bandit-PL
 ‘They brought the bandits to my father’s house, since he was the chief, the leading person.’
 (MAN: 174)

It was only with Ubryatova’s work on Dolgan that the habitual participle was recognised as the basis of a separate verb paradigm expressing the habitual ‘mood’ in either present or past tense, and since then this view has been widely accepted for both Dolgan and Sakha (Ubryatova, cited by Korkina (1970: 221)). While in the Russian literature on Dolgan and Sakha the habitual is defined as a modal category, I prefer for the remainder of this discussion to classify it as aspect instead, following Comrie’s definition that aspects are different ways of viewing the internal temporal constituency of a situation. An example for both present and past is given in 6.10 and 6.11.

DOLGAN

- (6.10) *On-tu-gun bieχ küł-e:čči-bin ile*
 that-DER-ACC.2SG always laugh-HAB-PRED.1SG really
 ‘I always really laugh at that.’
 (LKS: 227)
- (6.11) *ol kördük kihi köh-ö hıldz-a:ččī e-ti-bit*
 that similar human migrate-SIM.CV go-HAB be-PST-1PL
 ‘That is how we used to migrate.’
 (PPK: 40)

Judging from various grammars, most of the formal and functional properties of the participle seem to be shared between Dolgan and Sakha. However, the comparison also suggests certain differences. Formally, these include differences

in the allomorphy of the suffix, as well as differences in negation strategy. Functionally, the verbal use of the participle seems more widespread in Dolgan than in Sakha. In the remainder of this chapter I will focus on the following questions: a) do the patterns described in the grammars match with my own corpus data; b) are there more differences between Dolgan and Sakha than previously described; and c) what is the most probable scenario to explain these differences?

6.3.3 DIFFERENCES BETWEEN DOLGAN AND SAKHA

6.3.3.1 -A:ččI vs. -A:ččIk

As was mentioned above, Korkina's account of the verbal system in Sakha describes two variants of the habitual participle suffix, -A:ččI and -A:ččIk. Although -A:ččIk is mentioned in Ubryatova's description of Dolgan, it is doubtful to what extent this allomorph really is part of the Dolgan language. In my Dolgan corpus of spoken texts (16,250 words) there are 227 instances of the habitual participle, but there is not a single instance of -A:ččIk among them. Moreover, while this allomorph is recognised by Ubryatova, it is not mentioned in the later grammar of Dolgan by Artemyev (2001). Most importantly, on explicit inquiry about this suffix, Dolgan speakers say it is not part of their language. This gives the impression that Dolgan is different from Sakha in this respect. However, a fair comparison requires an investigation of the oral corpus for contemporary Sakha, and surprisingly this also did not yield any instances of -A:ččIk. Whether this result is due to a recent change in both languages, or whether -A:ččIk has always been marginal and its absence in both corpora is due to chance is impossible to determine without further detailed historical research. However, the data are sufficient to show that the grammars are not always a reliable guide to contemporary spoken language, and that ostensible differences between Dolgan and Sakha on paper may not prove significant upon closer investigation of spoken corpora.

6.3.3.2 NEGATION

According to the literature (e.g. Korkina 1970: 223) the habitual in Sakha can be negated by adding possessive person marking (agreeing with the subject) to the

habitual participle, followed by the unmarked negation noun *huoχ* (6.12). Alternatively this can be done by adding the invariable third person possessive marking to the participle, followed by the predicative person-marked negation noun *huoχ* agreeing with the subject (6.13):

SAKHA

(6.12) *Min bar-a:ččī-m huoχ*
 1SG go-HAB-POSS.1SG NEG

(6.13) *Min bar-a:ččī-ta huoχ-pun*
 1SG go-HAB-POSS.3SG NEG-PRED.1SG

'I usually don't go.'

(Korkina 1970: 223-24)

The same strategies are mentioned for Dolgan in Artemyev (2001: 201), but in the spoken corpora for both Dolgan and Sakha I only find instances of the second type. However, it is worth mentioning that the Sakha narratives displaying this negation construction were narrated by speakers of the Olenek district in the north of the Sakha Republic, which borders on the area where Dolgan is spoken. Therefore it is possible that only part of the linguistic variation in Sakha is reflected in the data, in particular the variants that are very similar to Dolgan. To confidently make a statement about the possible absence of the first negation strategy in Sakha, one would need a more complete picture of habitual negation in several dialects of Sakha, including the ones geographically remote from the Dolgan-speaking area.

6.3.3.3 VERBAL, NOMINAL AND ADJECTIVAL USE

According to the grammars, participles in Dolgan and Sakha may be verbal, nominal and adjectival in character, but the frequency of these usages is quite different across the languages. This has been noted previously; Ubryatova defines the Dolgan habitual participle primarily as the basis for the verbal paradigm of the habitual 'mood', with a possible usage as a noun or adjective, whereas in Sakha the nominal use is particularly frequent (Ubryatova 1985:182, 183). This implies that in Dolgan the verbal use is expected to be dominant, and in Sakha the nominal use. This pattern was confirmed and reinforced by current speakers of Dolgan, who

considered the nominal use not even grammatical (see example 6.17). It is of course in the nature of participles as ‘verbal nouns’ or as ‘nominal verbs’ to pose difficulties for a clear-cut categorisation into word classes on the basis of formal criteria, especially in those cases where they are not formally marked. For example, the unmarked participle can have a nominal (6.14), a verbal (6.15) or adjectival meaning (6.16), depending on the context as can be seen from the examples.

SAKHA

- (6.14) *īt-a:ččī* *bastiŋ-a* *e-ti-m (...)*
 shoot-HAB best-POSS.3SG be-PST-POSS.1SG
 ‘I was the best shooter of all.’ (AICH: 177)

DOLGAN

- (6.15) *Ol* *ih-en* *bar-an* *kel-en,* *bieχ* *kīrb-a:ččī (...)*
 PRT drink-SQ.CV go-SQ.CV come-SQ.CV always hit-HAB
 ‘When he came home after drinking he always beat me (...)’ (LKS: 165)

- (6.16) (...) *iti* *olor-o:čču* *oyo-lor-um* *barī-ta*
 this sit-HAB child-PL-POSS.1SG all-POSS.3SG
 taba *üöreg-iger* *bar-bīt-tara* (...)
 reindeer education-DAT.3SG go-PST.PTC-POSS.3PL
 ‘These sitting children all went to study for reindeer veterinarians.’ (PPK: 55)

This fuzziness of word classes, in particular for participles, may give the impression that a classification of habituals is too problematic to allow a sensible comparison between Dolgan and Sakha. However, in the majority of cases the ambiguity is easily resolved by context and/or certain formal and semantic criteria, such as presence of case marking, position in the clause, or agent/patient-like semantics. In this way it is possible to compare the frequencies of occurrence of these rather objective features, without forcing them into rigid word classes, which in reality may have fluid boundaries.

6.3.3.4 FREQUENCY OF USE

In order to compare the use of the habitual across Dolgan and Sakha, a Filemaker database was created, and all instances of the habitual participle in the spoken text corpora of Dolgan and Sakha were coded for the following formal and semantic properties: case marking, person marking (possessive or predicative), location of the person marking (on participle, auxiliary or negation word), person, tense, polarity and semantic function.

Participles that show case marking or are used in subject or object position in the sentence were identified as nominal. In theory, such forms could also have a predicative function in complement clauses, in which case they could be classified as verbal, but in my corpus I had no such instances and thus the classification is unambiguous. Participles without case or person marking that occur as modifiers of a noun were defined as adjectival. This is in accordance with the general shape of adjectives in Dolgan and Sakha, which never take agreement marking. Participles with predicative suffixes that occur in sentence-final position were classified as verbal. Potential ambiguity between unmarked nominals, adjectives and third person singular verbs, which are also unmarked, was resolved by context and the predominantly sentence-final position of verbs.

I compared the Dolgan and Sakha corpora with respect to the overall frequency of habitual participles in general, as well as their use as members of the individual word classes of verb, noun and adjective, respectively. The overall frequency was determined as the percentage of habituals over the total number of words in the corpus. This measure only makes sense because of the high comparability of the Dolgan and Sakha corpus with respect to text genre and mode of transcription. Both corpora consist of mostly life stories (as opposed to e.g. procedural texts or folk tales that could be different in vocabulary or structure), and the mode of transcription of the Dolgan texts was matched deliberately to the style used for Sakha in order to facilitate comparison. Therefore, this measure is justified.

The analysis shows that there are significant differences between the languages both in the overall frequency of habituals and in the word classes. The results are summarised in Tables 6.1 and 6.2 below.

Table 6.1: Frequency distribution for habitual participle in Dolgan and Sakha

Language	No of words in corpus	No. of habituals	% of total no. of words
DOLGAN	16,250	227	1.4
SAKHA	29,417	72	0.2

Table 6.1 shows that the overall proportion of habituals in the Dolgan corpus is considerably higher than in Sakha. In the Dolgan corpus 1.4% of all words (227 instances) are a habitual participle, whereas in Sakha this is only 0.2% (or 72 instances). A chi-square test for homogeneity of the two distributions is highly significant ($p < 0.0001$, $df = 1$), demonstrating that the difference between the proportions of habitual participles in the two languages is unlikely to have occurred by chance. This quantifies the statement made in the grammar of Dolgan that this participle is more common in Dolgan than in Sakha (Ubryatova 1985: 184).

Table 6.2 focuses on the occurrence of the habitual participle as part of different word classes, and here too the observed differences are highly unlikely to be due to chance alone.

Table 6.2: Comparison of habitual participle and its word class in Dolgan and Sakha

Language	No. of habituals	% Verb.	% Noun	% Adj.
DOLGAN	227	99.1	0	0.9
SAKHA	72	72.9	25.7	1.4

First, a chi-square test indicates that the category of ‘habitual participle’ across Dolgan and Sakha is significantly different, in other words, the distribution of verbal, nominal and adjectival use within this category is non-homogeneous across the two languages ($p < 0.0001$, $df = 2$). Further investigation into which factors are the cause of this significant difference confirms what can be read from Table 6.2 with the naked eye: a Fisher exact test comparing the different uses across the two languages shows that the verbal, as well as nominal, use of the participle is significantly different ($p < 0.0001$, $df = 1$). The verbal use in Dolgan is significantly higher, whereas the nominal use is significantly lower than in Sakha ($p < 0.0001$, $df = 1$). The difference in adjectival use is not significant ($p = 0.55$, $df = 1$).

These results match the statement that in Dolgan the verbal use of the habitual participle is very common (Ubryatova 1985: 184), whereas in Sakha it

conversation, the participle in *-A:ččī* is constructed here in a similar way to how it occurs spontaneously in Sakha texts, as was presented earlier and is repeated below in (6.19).

SAKHA

- (6.19) *dʒe mama-bar ha:maj tireχ buol-an*
 well mother-DAT the.most support AUX-SQ.CV
χa:l-l-īm, kömöölöh-ö:ččü (...) buol-an χa:l-l-īm.
 remain-PST-POSS.1SG help-HAB (...) AUX-SQ.CV remain-PST-POSS.1SG
 ‘Well I remained my mother’s biggest support, I remained her helper.’
 (ARR: 031)

Although this form does occur in written Dolgan texts, the more common way to express agent nouns in Dolgan speech is to use it attributively in combination with the word *kihi* ‘person’, or *oyo* ‘child’, by which it acquires a modifying rather than a substantival function⁷:

DOLGAN

- (6.20) *mas abirat-a:ččī oyo kel-bit*
 wood chop-HAB child come-PST.PTC
 ‘The boy who helped you with the wood has arrived.’ (Elicited)

While example 6.20 is intended to illustrate how agent nouns can be expressed in Dolgan, it also serves as a good example of the changing face of the participle. Depending on the interpretation, *abirat-a:ččī* ‘chop-HAB’ can either be read as an adjective modifying *oyo* ‘child’, leading to the translation ‘the wood-chopping child’, or as the predicate of a relative clause, as is reflected in the translation of the example above. Moreover, if this were a Sakha example, *oyo* ‘child’ could be omitted, giving *abirata:ččī* the interpretation of an agent noun meaning ‘wood chopper’. While this would change the meaning of the sentence in that it does not specify for the young age of the woodchopper, it is correct from a grammatical point of view.

⁷ It needs to be mentioned here that another common way to encode agent nouns is through the attachment of the suffix *-Sīt*, which is not a derived verb form but a proper agent nominaliser only used for this purpose, e.g. *taba-hīt* [reindeer -AG.NLZR] ‘reindeer herder’.

This example illustrates the idea introduced in section 6.3.3.3 that the boundaries between word classes in Dolgan and Sakha can be fluid. While some languages may have a clear-cut distinction between nouns, verbs and adjectives, examples like the above suggest that for Dolgan and Sakha this division may be a linguistic construct for analysis rather than a reality. Nevertheless it can be shown on the basis of the more objective criteria, such as case marking, position in the sentence, and potential for modification, that the nominal use of the habitual participle in Dolgan is very marginal.

Summarising, we can say that the habitual participle displays three main differences between Dolgan and Sakha: 1) an increase in its overall frequency, 2) an expansion of the verbal use within the aspectual verb paradigm 3) the disappearance of nominal use

6.4 PROBING THE CAUSE OF THESE DIFFERENCES BETWEEN DOLGAN AND SAKHA

6.4.1 LANGUAGE-INTERNAL MOTIVATIONS

Finding an explanation of these differences between Dolgan and Sakha, language-internal as well as language-external factors should be considered. With respect to the first, it is worth remembering the idea mentioned in section 6.1 that the habitual aspect and present tense are tightly interconnected. If this is true, one could imagine that the contiguity of these two grammatical categories led to a fading boundary between the domains of use of the habitual and non-habitual present tense, and that speakers of Dolgan/Sakha began to use the habitual form in a wider context. Instead of using forms in *-A:ččl* only with a clearly habitual meaning, they employed it also to describe less obviously habitual actions, for which Sakha would use the non-habitual present tense, thus extending its domain of application, and potentially its frequency. These semantics could then have spread to other tenses (e.g. past) as well.

Plausible as this language internal account may be, it leaves unexplained why the frequent use of *-A:ččl* remains restricted to the Taimyr Peninsula and why the participle is no longer used as an agent nominaliser. Since we know that Dolgan history is characterised by intense contact with Tungusic people and their languages, influence from these languages on the development of these features should be taken in to account as well.

6.4.2 LANGUAGE-EXTERNAL MOTIVATIONS

6.4.2.1 MORPHOSYNTACTIC PROPERTIES OF HABITUAL IN TUNGUSIC LANGUAGES

Like the Turkic language family, the Tungusic languages do not have one single way to encode habitual aspect, but display an array of morphological devices to express this category. According to the literature, there is no single reconstructable Tungusic suffix or structure expressing habitual aspect, from which the constructions in today's languages could all be derived. The only reconstructed form mentioned in Benzing (1956: 1067 (119)) is **-wā.čī*, which he labels as a marker of iterative aspect, and which is reflected in the North Tungusic languages Even and Negidal as *-WE:č* and *-vāč*, respectively, with the modified function of habitual aspect.

In Evenki, the language with which Dolgan has been in closest contact, habitual aspect can be expressed in two ways: by a participial construction employing the habitual participle in *-vki* and an auxiliary verb *bi-* 'to be' or by means of the suffix *-ηnA*. While both constructions are mentioned in the literature on Evenki (e.g. Nedjalkov 1997: 247, Bulatova & Grenoble 1999: 32, 40, Boldyrev 2007: 669-670) it is not clear from these sources what the difference in meaning or context of use between them actually is. For example, Boldyrev (2007: 669) describes the habitual aspect with *-ηnA* (in his words the 'present habitual tense')⁸ as reflecting "a repetitive, habitual, typical action, presented in the wider understanding of the present tense, and not connected with the moment of speech"⁹. He goes on to say that it "correlates with the habitual participle This participle ... represents the action in which the grammatical subject is involved as its characteristic, and is normally expressed predicatively"¹⁰. Thus it seems that both habitual structures share the property of representing an action as 'typical' or 'characteristic' of the grammatical subject, regardless of tense or moment of speech, which is a fairly common meaning of habitual aspect cross-linguistically (Dahl 1985: 100). Nedjalkov does not specify a difference in meaning either, except

⁸ Болдырев (2007: 669): "Настоящее обычное время." (translation mine).

⁹ Болдырев (2007: 669): "Это время выражает повторяющееся, обычное, типичное действие, представленное в широком плане настоящего времени, не связанного с моментом речи." (translation mine).

¹⁰ Болдырев (2007: 669-670): "Настоящее обычное время коррелирует с причастием обычным (...). Это причастие (...) обозначает присущее грамматическому субъекту действие как признак (...) выражаемый обычно предикативно." (translation mine).

that he adds the possible reading of *-vki* constructions as ‘potential’ and ‘probable’ (Nedjalkov 1997: resp. 236, 266). However, from a corpus of Evenki folklore texts it appears that the habitual in *-ŋnA* also has connections with the potential reading, especially the negative potential. Neither Boldyrev nor Nedjalkov make explicit what the ‘correlation’ between the two habitual constructions involves exactly, and their comparison is too brief to confidently disregard the possibility of any differences in meaning or in pragmatic use. However, since data from the grammars are currently the only available source, I will assume for now that semantic differences play a minor role and that the difference between the two habituals is primarily a matter of form and dialect choice.

The Evenki habitual in *-ŋnA* is an aspectual suffix that can be attached to any verb stem and is followed by tense and person marking:

EVENKI

- (6.21) *bu* *enin-du-ver* ***bele-ŋne-re-v***
 we mother-DAT-REFL.POSS help-HAB-NONFUT-1PL.EXCL
 ‘We usually help our mother.’ (adapted from Nedjalkov 1997: 247)

In contrast, the habitual aspect formed with *-vki* is an analytical construction formed with the participle, and optionally followed by a form of the auxiliary *bi-* ‘to be’ to form a full predicate.

EVENKI

- (6.22) *Nungan* ***tangi-vki*** ***bi-si-n***
 she read-HAB be-PRS-3SG
 ‘She usually reads books.’ (Nedjalkov 1997: 236)

As was mentioned above, examples like 6.22 can also have the potential reading ‘she can read’ and what Nedjalkov calls the ‘universal tense’, which could correspond to what Boldyrev calls ‘tenseless’, indicating that the focus is not on the moment of speech, but rather on the habituality of the action, which, as we have just seen, was said to be the meaning encoded by the suffix *-ŋnA*. Neither construction matches the morphological structure of the habitual in Dolgan perfectly. The *-vki* construction matches Dolgan in that it is a habitual participle, but differs in that it is exclusively analytical (see example 6.22), whereas the habitual in Dolgan is formed synthetically, at least in the present tense (see example 6.10). The mirror image applies to the *-ŋnA* construction, which is

synthetic, like Dolgan, but which is not a participle. With respect to the nominal use of the habitual participle, this is possible in Evenki, but rare (see section 6.4.2.2).

In Even there are two suffixes encoding the habitual aspect, *-WE:č* and *-Gr(E)* (Malchukov 1995: 15, Benzing 1955: 42, 43). Both suffixes are attached to the verb stem, and can be followed by tense and person markers, as in examples 6.23 and 6.24 (Pakendorf, fielddata).

EVEN

- (6.23) *Eńeńe te:leŋ-e-d-dzo:t-te-n ti:tel bi-si-ten*
 grandmother tell-EP-PROG-GNR-NONFUT-3SG long.ago be-PST-POSS.3PL
oŋoč-i-l-dula-da-ka bej-u dzeb-mege-r.
 Oroch-EP-PL-LOC=PRT=EMPH man-ACC eat-NLZR-PL
 ‘My grandmother used to tell that a long time ago amongst the Evens there were cannibals.’ (EPA: 003)

- (6.24) *Bi dʒugani-du čumrobotnitsaj gurgewči-wre-re-m (...)*
 1SG summer-DAT yurt.worker.INST.R work-HAB-NONFUT-1SG (...)
 ‘In summer I work as a yurt worker (...)’ (RDA: 019)

The meaning of *-WE:č* is described in Cincius (1952) as “... an action which is carried out habitually under certain conditions”¹¹. Malchukov, on the other hand, classifies the suffix as iterative, but with an “usitative-habitual meaning” (Malchukov 1995: 15). Cincius’ description of *-Gr(E)* is that “the action was carried out not once”¹², which is confirmed by Malchukov, who adds that it primarily refers to events in the past (Malchukov 1995: 15). However, recent findings seem to indicate that the difference between the two suffixes is not so much in their semantics as it is in their geographical distribution (Pakendorf, pers. comm.): *-WE:č* is very common in the Even dialect of Kamchatka, whereas *-Gr(E)* is most commonly found in the western dialect of Sebjan-Küöl.

Negidal also employs two strategies to express habitual aspect, however, they differ from each other in structure. On the one hand, there is the aspectual suffix *-vāč*, related to *-WE:č* in Even, and on the other hand, there is a participle

¹¹ Цинциус (1952: 742): “Означает, что речь идёт о действии, обычно совершающемся при тех или иных условиях.”

¹² Цинциус (1952: 742): “Означает, что действие совершалось не раз.”

The southern Tungusic languages also show variation when it comes to the encoding of habitual aspect. In Udighe, the habitual is expressed only analytically by the impersonal present participle followed by the auxiliary *bi-* ‘to be’, which takes the person and tense marking. In the present tense this auxiliary verb is optional and is “typically (...) omitted” (Nikolaeva & Tolskaya 2001: 218).

(6.25) *nua-ni* ***wakca-i***
 he-3SG hunt-PRS.PTC
 ‘He usually hunts.’ (Nikolaeva & Tolskaya 2001: 218)

UDIGHE

(6.26) *nua-ni* **wakca-i** **bie**
 he-3SG hunt-PRS.PTC be.PRS.HAB
 ‘He usually hunts.’ (adapted from Nikolaeva & Tolskaya: 218)

UDIGHE

(6.27) *wakca-i* *bi-si-mi*
 hunt-PRS.PTC be-PST-1SG
 ‘I used to hunt.’ (Nikolaeva & Tolskaya: 219)

(6.28) *ag'a zugdii wo-isi-ni bu belesi-u bi-s'e-u*
 brother house.RFL make-PC-3SG we help-1PL.EXCL be-PF-1PL.EXCL
 'When my brother built a house, we helped him.' (Nikolaeva & Tolskaya: 219)

In Nanai, the aspect of “duration, multiplication and permanency”¹³, is expressed by the suffixes *-či*, *-vači*, *-si*, and *-so*, the choice of which seems to be determined by semantic properties of the verb stem (Avrorin 1961: 45-46). Of these suffixes, *-vači* and its phonological variant *-veči* can be recognised from the North Tungusic languages Even and Negidal and from Avrorin’s description it seems that this suffix occurs predominantly with verb stems denoting motion, as in *iveči* [*i-veči*-, enter-HAB ‘to enter often’]. Although verb stems with which the suffixes occur may be put in very rough semantic categories, semantic differences between these suffixes themselves are not specified in the description of Nanai. Furthermore, Nanai employs the impersonal present/future participle, extended with the suffix *-ni*, to express the “habituality of an action” (Avrorin 1961: 91)

NANAI

- (6.29) *adim-ba* *kiutel-di* ***va-o-ri-ni***
 beluga-ACC hook-INST kill-?-PRS.PTC-HAB
 ‘They catch beluga with a fishhook.’ (Avrorin 1961: 91, glossing mine)

This overview has shown that a formally marked category of habitual aspect is common in the Tungusic language family, but that the strategies to express this grammatical category are not homogeneous. An overview of the strategies discussed above is given in Table 6.3 below.

While the synthetic Proto-Tungusic **-wā.či* is reflected in Even, Negidal and Nanai, an additional analytical strategy is used in Evenki, Negidal and Nanai. Evenki also displays a synthetic strategy that is not related to the Proto-Tungusic suffix. The choice of strategy does not seem to correlate with a particular branch (northern or southern) of the Tungusic language family.

¹³ Avrorin (1961: 45): “Вид длительности, многократности и постоянства.”

Table 6.3: Encoding of habitual events in a number of Tungusic languages

Branch	Language	Analytic	Synthetic
Proto-Tungusic			*-wā.č̣i
North Tungusic	Even		-WE:č̣
			-Gr(E)
	Negidal	HAB.PTC -vki + bi 'to be'	-vāč̣
	Evenki	HAB.PTC -vki + bi- 'to be'	-ŋna-
South Tungusic	Udighe	PRS.PTC -i + bi- 'to be'	
	Nanai	PRS/FUT.PTC -ni	-vač̣i
			-č̣i
			-si
			-so

To tie this back to the observed differences between Dolgan and Sakha, it can be concluded that there is no direct match between Dolgan and the Tungusic languages as far as the construction is concerned. However, they do behave identically with respect to the purely verbal use of the participle. While in Sakha the habitual participle is used as an agent nominaliser, this is not possible in Dolgan and the Tungusic languages Evenki and Negidal.

6.4.2.2 FREQUENCY OF HABITUAL ASPECT IN TUNGUSIC LANGUAGES

Even and Udighe

For the comparison of frequencies of habituals between Dolgan and Tungusic languages the best available source was a corpus of Even texts. This is the only Tungusic language for which a spoken text corpus is available and which is comparable in size, text genre and annotation style to the corpora used for Dolgan and Sakha. Although Even and Evenki are different languages, they are closely related to each other, and therefore an investigation of this language may still provide useful insights with respect to the use of habituals. The corpus (collected and transcribed by Brigitte Pakendorf) contains spoken texts from two Even dialects, the eastern Even dialect of Kamchatka and the western dialect of Sebjan Küöl. For both dialects the two habitual suffixes -WE:č̣ and -Gr(E) were counted. The results are presented in Table 6.4.

Table 6.4. Frequency distribution of habituals in two dialects of Even

Dialect	No. of words	-WE:Č		-Gr(E)		Total	
		Nr.	%	Nr.	%	Nr.	%
Kamchatka	21,700	521	2.4	11	0.05	532	2.5
Sebjan Küöl	42,000	216	0.5	1000	2.4	1216	2.9

The frequency of the two suffixes is indicated in absolute numbers as well as in percentage of the total number of words in the corpus. Apart from the fact that the use of the two suffixes is clearly geographically conditioned (-WE:Č is found in Kamchatka and -Gr(E) almost exclusively in Sebjan Küöl¹⁴), the overall frequency of habituals in the two dialects is comparable (2.5% in Kamchatka and 2.9% in Sebjan Küöl), but much higher than in Dolgan (1.4%) and Sakha (0.2%).

A very preliminary survey of Udighe texts (Southern Tungusic) shows a similar pattern. The mini-corpus used for this impressionistic overview is based on glossed texts at the back of the descriptive grammar of Udighe (Nikolaeva & Tolskaya 2001). In this collection of 1,700 words, 47 instances of habitual aspect were encountered, which is 2.8% of the total number of words. While the corpus is very small and the reliability of these numbers is questionable as a representation of the language as a whole, the similarity of this percentage to the data from Even is striking, and may indicate that in the Tungusic family a frequent usage of habituals is common.

Evenki

For Evenki no spoken text corpus was available at the time this research was conducted, which is equivalent to Dolgan, Sakha and Even in size and genre. However, as a best alternative, a survey of Evenki folklore texts was used to get a first impression. The study of this material revealed an interesting diversity in the use of habituals across different Evenki dialects and it is worth discussing the findings in this context. It needs to be emphasised that the findings cannot be

¹⁴ Although -WE:Č is found in the dialect of Sebjan Küöl, it should be mentioned that its use is lexically determined, and is restricted to certain verbs, especially *bi-* 'be', whereas -Gr(E) can occur with any verb (Pakendorf, pers. comm.).

more than impressionistic due to their nature, and the outcomes need to be investigated in more detail once an annotated spoken corpus becomes available. The hypotheses formulated on the basis of this survey are primarily postulated to define useful directions for future research rather than to answer questions with certainty.

The corpus (18,247 words) contained folklore texts from three different dialects of Evenki, the Symiskij and Ilimpijskij dialects and the dialect of Podkamennaya Tunguska. The Ilimpijskij dialect borders Dolgan geographically, while the dialects of Podkamennaya Tunguska and Symiskij are spoken further south and west from the Dolgan-speaking area. The texts were coded for the habitual constructions introduced in section 6.4.2.1, the participle construction in *-vki* and the synthetic form in *-ŋA*¹⁵. For all three dialects, I computed the frequency of their occurrence as a percentage of the total number of words, the results of which are summarised in Table 6.5.

Table 6.5. Overall frequency of habitual mood in Evenki

Dialect	No. of words	No. of habituals	% of all words
Ilimp.	1,501	26	1.73
PKT	5,746	46	0.80
Syms	8,000	51	0.64

The first observation from Table 6.5 is that Evenki is not a homogeneous unit when it comes to the frequency of habituals across the different dialects. This impression is confirmed by a chi-square test comparing the three distributions ($p < 0.0001$, $df = 2$). The Ilimpijskij dialect shows the highest ratio of habituals taken over the total number of words (1.73%), followed at some distance by the dialects of Podkamennaya Tunguska (0.8%) and the Symiskij dialect (0.64%). A Fisher exact test comparing the frequencies for every possible pair of dialects identifies the Ilimpijskij dialect as the cause of this non-homogeneous picture. The Ilimpijskij dialect is significantly different from the other two dialects (Ilimp. vs. PKT: $p = 0.003$, Ilimp. vs. Syms.: $p < 0.00011$), whereas the frequency difference between the dialect of Podkamennaya Tunguska and the Symiskij dialect is not significant ($p = 0.30$).

¹⁵ Thanks to Jana Neuwirt for the coding work.

Comparison of the Evenki dialects with Dolgan and Sakha shows that all language pairs are significantly different in their use of habituais, except the Ilimpijskij dialect of Evenki and its geographical neighbour Dolgan. This outcome is represented in Table 6.6, in which the boldly printed p-value of 0.31 is not significant.

Table 6.6. P-values for comparison of habitual frequency between Evenki dialects, Dolgan and Sakha

	Ilimpijskij	PKT	Syms
Dolgan	p = 0.31	p = 0.0004	p < 0.0001
Sakha	p < 0.0001	p < 0.0001	p < 0.0001

The impression of similarity between the Ilimpijskij dialect of Evenki and Dolgan is strengthened when we look at the strategy that is used to express habitual aspect. In contrast to the other two Evenki dialects, the Ilimpijskij dialect uses predominantly the habitual participle to express habitual meaning, and as we know this is also the case in Dolgan. Table 6.7 summarises the distribution of the two strategies in the Evenki dialects.

Table 6.7. Frequency of habitual on -vki and -ŋnA in Evenki dialects

Dialect	No. of words	-ŋnA	% of all words	-vki	% of all words
Ilimp.	1,501	5	0.33	21	1.38
PKT	5,746	19	0.33	27	0.45
Syms	8,000	50	0.63	1	0.01

With respect to the suffix -ŋnA there is no significant difference in frequency of use across the three dialects, The construction with -vki establishes variation ranging from virtually absent in the Symskij dialect (0.01%) to 0.45% in the dialect of Podkamennaya Tunguska and a significantly different percentage of 1.38% in the Ilimpijskij dialect (p < 0.001 for a Fisher exact test).

Apart from frequency, the habitual participle in -vki shows similarity to the Dolgan use of the participle in -A:ččī in other respects. While it can occur with attributive, nominal and predicative function, as exemplified in sentences 6.30, 6.31, 6.32, Nedjalkov notes in his grammar that the overwhelming majority of cases is predicative (adapted from Nedjalkov 1997: 268).

EVENKI

- (6.30) *Havali-vki beje suru-re-n.*
 work-HAB man go.away-NONFUT-3SG
 ‘The man who usually works went away.’

- (6.31) *Aičimni bumu-d’e-vki-ve aj-ra-n.*
 doctor be.ill-IMPV-HAB-DEF.ACC cure-NONFUT-3SG
 ‘The doctor healed the person who was often ill.’

- (6.32) *Nuġan tangi-vki bi-si-n*
 he read-HAB be-PRS.3SG
 ‘He usually reads’ or ‘He can read.’ (adapted from Nedjalkov 1997: 236)

In the same grammar, he mentions that the attributive participle has a very restricted use (Nedjalkov 1997: 276), and as a noun, the habitual participle does not occur as an agent nominaliser, which is the main nominal function of the habitual in Sakha, but is of questionable status in Dolgan (see Section 6.3.3.3). This is confirmed by the results from the Evenki text corpus, where *-vki* occurs only with a verbal function. Thus, the predominantly verbal employment of the participle in the Ilimpijskij dialect corresponds to the way it is used in Dolgan.

The fact that the Ilimpijskij dialect of Evenki behaves significantly different from its genealogically related neighbouring dialects, in combination with the similarity in frequency to its unrelated neighbour Dolgan, makes the idea that the deviating pattern in Evenki is due to contact tempting. However, the data have shown that both the Ilimpijskij dialect of Evenki and Dolgan deviate from their closest relatives. If the abovementioned similarity was motivated by contact, then linguistic data alone are not enough to establish the source and recipient language, and thus the direction of change.

6.5 INTERPRETATION AND CONCLUSION

Thus the data presented here have sketched the following picture. Dolgan and Sakha differ in their use of habitual aspect in overall frequency (1.4% in Dolgan vs. 0.2% in Sakha) and with respect to its character. In Dolgan the habitual participle is used almost exclusively in a verbal way, while in Sakha the nominal use is also

relatively common. The nominal use in Dolgan is so rare in spontaneous speech that the nominal status of the participle can be questioned. At the same time, the Evenki data also display dialectal variation, whereby the Ilimpijskij dialect differs significantly from the other two, and shows an overall frequency of habituals that is comparable to Dolgan. To facilitate interpretation, the frequencies for Evenki as well as for the other dialects and languages are repeated in the table below.

Table 6.8. Frequency of habituals per dialect and per language

Family	Language	Dialect	% of HAB/ dialect	% of HAB/language
Turkic	Sakha		0.2	0.2
	Dolgan		1.4	1.4
Tungusic	Evenki	Ilimp.	1.7	0.8
		PKT	0.8	
		Syms.	0.6	
	Even	Kamch.	2.5	2.7
		SK	2.9	
	Udighe		2.8	2.8

This similarity between two unrelated languages, Dolgan and the Ilimpijskij dialect of Evenki, which at the same time differ from their respective sister languages, leads to the idea that the observed similarity in the frequent use of the habitual aspect could be an areal feature motivated by contact. Given the history of the Dolgans and Evenks on the Taimyr Peninsula, this is certainly not unthinkable.

However, even if contact did play a role, the question remains which language accommodated to which? As we have seen, the linguistic data alone are not sufficient to answer this question. The high frequency of habitual aspect in Tungusic languages and the low frequency in Sakha stimulates the thought that Dolgan adapted to the Tungusic pattern. However, the data from Evenki complicate this picture. The divergence in frequency across the Evenki dialects makes it hard to determine which of the dialect frequencies represents 'typical Evenki', if there is such a thing. The average frequency of habituals varies from 0.6% to 1.7% and it is unclear which proportion best represents the language use of the average Evenki speaker.

On the one hand, the high frequency observed in the Ilimpijskij dialect could be a remnant of a typical Tungusic pattern, which would be compatible with the high frequencies recorded for Even and Udighe. In that case, the high percentage

in Dolgan could have emerged as an accommodation to Evenki, most probably brought about by Evenki speakers who shifted to Dolgan/Sakha. The low frequency in the other two dialects could be attributed to contact of these dialects with Turkic languages, in particular Sakha, where use of habitual is low. On the other hand, the low frequency in the Symskij dialect and the dialect of Podkamennaya Tunguska could represent the Evenki standard. In this case the speakers of the Ilimpijskij dialect would have accommodated to Dolgan, in which the frequent use of habituals would have developed language-internally.

Despite all caveats, the former scenario seems more likely for a number of reasons. First, at the moment we have no plausible justification for a purely language-internal account. Support for such an account would come from historical information about Dolgan, which would allow us to track the use of the habitual participle through time. Alternatively, a language-internal account would be attractive if a similar phenomenon were observed in other Sakha dialects outside of the contact area, or if the neighbouring languages did not have habitual aspect. However, none of these conditions apply to Dolgan. We have no historical material from Dolgan older than from the 1920's, which is long after the period of intense contact with the Evenks, the frequent occurrence of habituals is only observed in Dolgan, and habitual aspect is a prominent category in Tungusic languages, including Evenki. Therefore, the possibility of language external-motivation needs to be taken seriously.

Second, historical records mention Evenks shifting to Dolgan, rather than the other way round (although the opposite direction did of course occasionally occur as well). Finally, within language contact theory the transfer of frequency patterns is associated primarily with situations of language shift. While this is not an indisputable law, this tendency supports the idea that in the case of contact-induced change, speakers of Dolgan accommodated to Evenki rather than the other way round. More specifically, the described phenomenon could be classified as *frequential copying* (Johanson 1992: 175, 2002a: 13, 109, 2002b: 292) which means that

frequency patterns peculiar to model code units [source language units in the terminology used in this thesis, E.S.] are copied onto units of the basic code [or recipient language, E.S.] so that the latter undergo an increase or a decrease in frequency of occurrence. (Johanson 2002b: 292)

In Heine and Kuteva's words, "increased frequency of use is the driving force in establishing new use patterns" (Heine & Kuteva 2005: 47), during which a 'minor use pattern' in the recipient language expands to become a 'major use pattern' due to contact with a source language, which is a common phenomenon in contact situations cross-linguistically (Heine & Kuteva 2005: 44).

But even if we accept Evenki as the source language and Dolgan as the recipient language for the transfer of this frequency pattern, there is still the question with respect to the underlying process of the transfer. Dolgan speakers may have adopted the pattern of a commonly-used habitual into their language through the process of borrowing, or it may have appeared as a result of imposition by Evenki speakers who shifted to Dolgan and projected this pattern onto their target language. Since copying of morphosyntactic patterns is mostly associated with scenarios of language shift and imposition (Thomason & Kaufman 1988: 50, Van Coetsem 2000: 58, 59), and since we know from historical and genetic sources (Dolgikh 1963, Whitten et al. in preparation) that there must have been considerable number of Evenks who intermarried with the Dolgans and adopted their language (see Sections 2.3.2.3 and 2.6), the most realistic assumption is that the increase in use of habituals was introduced into Dolgan by speakers of Evenki who shifted to Dolgan. It is possible to imagine that a significant number of Evenki-speaking people, who were learning Dolgan, found in the Dolgan suffix *-A:ččI* a semantic equivalent to their own habitual suffixes *-vki* and *-ɣna*. Imperfect learning by the first generation and the presumed large number of Evenks that mixed and intermarried with the Dolgans, may have provided an excellent context for the new pattern to spread. According to Hickey (2010), this scenario is rather common in situations of language shift. He writes that

[w]hen shifting to another language, temporarily or permanently, adults expect the same grammatical distinctions in the target which they know from their native language. To this end they search for equivalents in the target language to categories they are familiar with. This process is an unconscious one and persists even with speakers who have considerable target language proficiency. If the categories of the outset language are semantically motivated then the search to find an equivalent in the target is all the more obvious. (Hickey 2010: 155)

The fact that contact with Mongolic had already led to the marginal use of the participle in *-A:ččI* as a verb in other Turkic languages, including Sakha before it spread to the far north, may have facilitated this process.

As an alternative explanation one could consider the possibility that neither Dolgan nor Evenki was the source of this change, but that it happened as a result of contact with an external third language, with which both Dolgan and Evenki were in contact. The most obvious candidates for such a scenario would be the Samoyedic languages Nganasan or Enets, which are spoken in the area as well. However, this possibility must be dispelled. First, although these Samoyedic languages have aspectual suffixes to express iterativity or durativity, they do not have a specific category for habitual aspect. Second, although the Samoyedic people have inhabited the Taimyr Peninsula the longest and their languages must have been widespread in the area, they lived relatively isolated from other ethnolinguistic groups, and interaction with other groups was infrequent when compared to the life-style of the trading Dolgans (see Chapter 2). Thus the relatively low intensity of contact, the absence of sociolinguistic dominance and the non-prominent use of habituais make the possibility of Nganasan as the source language of the changes in Dolgan and Evenki very implausible.

To summarise, it has been shown that Dolgan and Sakha differ significantly with respect to the use of the habitual participle. This applies to its overall frequency as well as to its use as an agent nominaliser. On the basis of the available data it is not possible to explain this difference conclusively, either as language-internal or as motivated by contact. However, the frequent use of habituais in Tungusic languages, the similarity between Dolgan and the Ilimpijskij dialect of Evenki and the history of contact between Dolgans and Evenks on the Taimyr foreshadow fruitful research in this domain for the future.

For this, more annotated text corpora are needed, in particular for spoken Evenki and for other Turkic languages. In addition, more detailed semantic analysis of the exact connotations and contexts of use for the different habitual suffixes in Evenki, as well as across Dolgan and Evenki would be helpful to determine the degree of overlap between the two languages. Nonetheless, the data enable us to quantify differences that have been mentioned anecdotally in descriptions of Dolgan and Sakha, and the first impressionistic results from this comparative corpus study generate hypotheses as to the cause of these differences. Finally, this study proves the importance of corpus studies in the investigation of synchronic variation and diachronic change, since they provide crucial information that cannot be acquired from grammars alone.

7.1. INTRODUCTION

An additional domain in which Dolgan differs from Sakha is word order. Compared to Sakha, Dolgan shows greater variation in the arrangement of constituents in transitive sentences, allowing more freely for orders other than the standard Turkic (and Sakha) SOV, in particular for SVO. This is not to say that Turkic languages show no variation at all and that Dolgan is the exception within the language family. Many Turkic languages do allow for variation, usually associated with particular discourse pragmatic functions such as topicalisation. To give an example, regardless of the fact that SOV is its pragmatically least marked order, some scholars claim that Turkish is essentially a language with free word order (Kornfilt 1997: 91). In addition to language-internal reasons for non-SOV constituent order, many Turkic languages spoken in the vicinity of languages belonging to other families (such as Slavic), have acquired greater flexibility due to contact with their neighbours. For example, word order in Khakas has become more flexible under the influence of Russian (Anderson 1998: 71), and in the West Rumelian dialects of Turkish spoken in Macedonia, SVO has become the unmarked word order under the influence of Macedonian (Friedman 2003: 66).

Rather than treating word order change by itself as an exotic phenomenon, the issues of interest for the current study are a) an investigation of the difference

in word order variation between Dolgan and Sakha; and b) how this difference can be explained. First I will show on the basis of quantitative analysis of word order patterns that the higher degree of flexibility in Dolgan is very unlikely to be due to chance. It cannot be attributed to the idiolect of certain individuals, nor does it correlate with a certain text genre or age category. I take this as evidence that this tendency is pervasive throughout the entire language, and that the present variation could eventually become established as a change.

After a review of some of the main ideas on word order change in the literature, I will argue that the variation in Dolgan word order is the result of Russian influence brought into the language by bilingual speakers of Dolgan. To substantiate this claim the sociolinguistic situation in the different Dolgan communities will be discussed and I will postulate that while this change is ongoing, the change is best explained in terms of two underlying processes of contact-induced change, depending on the linguistic dominance of the speaker: imposition in Dolgan people whose dominant language is Russian (i.e. typically the younger generations), and borrowing in those people whose dominant language is Dolgan, but whose way of speaking is influenced by the constant exposure to Russian.

7.2. WORD ORDER IN TURKIC LANGUAGES

Since the structure of Turkic languages is predominantly head-final, it follows that the unmarked word order in most languages is SOV. Within this statement, O needs to be understood as any kind of object, and V as any kind of predicate rather than only as a direct object and a verb, for which these abbreviations are normally used. This applies to finite (7.1, 7.2, 7.3) as well as to non-finite (7.4) clauses and is observed particularly strictly in the latter category (Johanson 1998: 57). In the following examples the object is marked in bold so the different clause orders can be spotted more easily:

TURKISH

- (7.1) *Hasan **kitab-ı** oku-du*
 Hasan book-ACC read-PST

'Hasan read the book.'

(Kornfilt 1997: 89)

KHAKAS

- (7.2) *min tayda paba-zina pu kniga-ni pir-e-m*
 I tomorrow father-3.DAT this book-ACC give-FUT-1

‘Tomorrow I will give this book to his father.’

(Anderson 1998:72, morpheme breaks mine)

UYGHUR

- (7.3) *saen suet ich-t-ing*
 You milk drank-PST-2.SG

‘You drank milk.’

(De Jong 2007: 101, glossing mine)

UZBEK

- (7.4) *Āybek-ni bu kitāb-ni yāz-yāni-ni bilā-mān*
 Aybek-GEN this book-ACC write-CV-ACC know-1SG

‘I know that Aybek has written this book.’

(Johanson 1998: 60, glossing mine)

However, in most Turkic languages variation in word order is not uncommon. Typically, a non-standard arrangement of constituents correlates with certain discourse-pragmatic functions. Constituents in sentence initial position normally have the interpretation of topic, whereas the focused element is found directly before the predicate (Johanson 1998: 58-59). In a pragmatically unmarked sentence these positions roughly correlate with the grammatical functions of subject and object, but this pattern may be reversed when other constituents are assigned the function of topic or focus, as in example 7.5. In this example, the sentence-initial position is occupied by the object *istakozu* ‘lobster’ instead of the subject *Hasan*, because the lobster is the topic of this sentence.

TURKISH

- (7.5) *istakoz-u Hasan Ali-ye ver-di*
 lobster-ACC Hasan Ali-DAT give-PST

‘(Speaking of) the lobster, Hasan gave (it) to Ali.’

(Kornfilt 1997: 200)

An additional ‘postpredicative position’ in Turkic languages, which is not included in the description of unmarked SOV sentences, is reserved for information that is not new, such as already activated topics, defocused constituents, or afterthoughts (Johanson 1998: 58). This is illustrated with an example from Turkish in 7.6, where

the sentence-final subject ‘Hasan’ represents shared background information. Kornfilt explicitly says that in this language the constituent in post-predicative position does not represent afterthoughts, but rather encodes shared knowledge or ‘backgrounding’ (Kornfilt 1997: 206), which is only compatible with the first two functions (topic and defocused constituent) described for Turkic by Johanson. However, the descriptions for Turkic and Turkish have enough in common to illustrate the function of postverbal slot with a sentence from Turkish.

TURKISH

- (7.6) *Ali-ye kitab-ı ver-di Hasan*
 Ali-DAT book-ACC give-PST Hasan
 ‘He gave the book to Ali, Hasan.’ (Kornfilt 1997: 206)

7.3. WORD ORDER IN SAKHA

In Sakha standard word order and its possible variants closely resemble the general Turkic pattern described above. Sakha typically employs the standard Turkic SOV order for unmarked transitive clauses (Stachowski & Menz 1998), but very often only O and V are overtly expressed due to the fact that Sakha is a pro-drop language, as can be seen from example 7.7. Full SOV sentences, in which all three core constituents are overtly expressed, are in fact very rare in spontaneous narratives (only 0.8% of all counted transitive clauses in the corpus).

SAKHA

- (7.7) *Bu Uolba hir-itten sü:rbe toyus kihi-ni*
 This Uolba place-ABL.3SG twenty nine person-ACC
ilđi-bit-tere, bu kirakij bayajı deriebine-tten
 take.away-PST.PTC-POSS.3PL this tiny INTNS village-ABL
 ‘From Uolba they took twenty nine people, from this very tiny village.’
 (ARR: 022)

This sentence seems to be neutral with regard to the relation between the object *sü:rbe toyus kihini* ‘twenty nine people’ and the verb *ilđi-bit-tere* ‘they took’, while the topic *Uolba* is placed in sentence initial position, and further specified as *bu kirakij bayajı deriebinetten* ‘from this little village’ as an elaboration and afterthought.

As in other Turkic languages, deviation from this basic word order pattern occurs for discourse-pragmatic reasons such as topicalisation, in which case the topic is fronted to clause initial position (Stachowski & Menz 1998). This is in agreement with data from the spoken corpus of Sakha. Within a set of 176 transitive clauses with overt expression of O and V (for details see Section 7.5.1) only 3 VO clauses are found (1.7%), reflecting the dominance of the OV pattern convincingly. In addition, 2 instances of OSV order are attested. All sentences with non-SOV word order have clear pragmatically marked connotations. Evidence for this is most clearly seen, or rather heard, in intonation patterns. In unmarked statements, sentence stress in Sakha normally comes on the final constituent. Since this is typically the verb, as a consequence of Sakha's SOV word order, in the average statement verbs are lightly stressed. Despite this being the unmarked prosodic pattern, sentences in which the final verb is in focus are still clearly distinguishable. In these cases the stress on the final verb is noticeably increased, and in addition the object can be moved to sentence-initial position to underline its topicality. This is exemplified in example 7.8, which displays OSV order. The sound recording reveals an unmistakable increase in stress on the verb *körbütüm* 'I have seen', showing that the focus of the sentence is the act of 'seeing', whereas the 'husband' in clause initial position fulfills the function of topic. This is further supported by the discourse context in which this sentence was produced. It is a story about a wedding, in which the participation of a wife (who is also the narrator) and a husband is typically presupposed. A third participant, who was ill and could therefore not come to meet the husband at the wedding, then said that he would not be able to meet the husband now, but that he has *seen* that husband before.

SAKHA

- (7.8) *En kergeŋ-ŋin min kör-büt-üm dir.*
 2SG spouse-ACC.2SG 1SG see-PST.PTC-POSS.1SG say.PRS.PTC
 'I've *seen* your husband, he said.' (ARR: 273)

In example 7.9, clause order is VO, and the sound recording shows a clear break between the verb *emti:r* 'he treats' and the object *tugu barit̃in*, suggesting that the object 'what, everything' is produced as an afterthought.

(7.9) *Em-ti:r* **tugu** **bari-tin,** *telepatija,*
 medicine-VBLZR.PRSPT what.ACC all-ACC.3SG telepathy
vse takoe.
 all.R such.R
 ‘He treats what, everything, telepathy and all that.’ (ARR: 256)

VO sentences for which the discourse-pragmatic function is undisputedly unmarked are not found in Sakha. From these data we must conclude that the unmarked word order pattern in Sakha is in line with the Turkic languages in general: transitive clauses are rather strictly SOV, and exceptions to this pattern occur only for particular discourse-pragmatic reasons.

In Dolgan, unmarked word order is also predominantly SOV, as is illustrated in examples 7.10 and 7.11. These sentences are neutral descriptions of what is usually done in preparation for migration (7.10) and how reindeer hides are prepared (7.11). They have a neutral intonation pattern in which none of the arguments is particularly stressed except for the light clause-final stress that, as in Sakha, characterises the unmarked prosodic pattern. Examples 7.10 and 7.11 also illustrate that Dolgan, like Sakha, is a pro-drop language, in which the S is frequently not overtly expressed within the clause.

(7.10) *taba* *tut-a-bīt* *buo* *oččoyo* *buollayina*
 reindeer hold-SIM.CV-1PL PRT then PRT
bolox-putugar *ayīs* *taba-nī* *kölüj-e-bit*
 balok-DAT.1PL eight reindeer-ACC harness-SIM.CV-1PL
 ‘We catch reindeer, and then for our balok, we harness eight reindeer.’
 (IMA: 10)

- (7.11) *taha:ra giniler-iñ mañnaj iti tiri: üle-tin*
 outside 3.PL-POSS.2SG first this skin work-ACC.3SG
üle-li:-ler
 work-VBLZR.SIM.CV-PRED.3PL
 ‘Outside they first do the work with the skin.’ (ESB: 04)

However, data from the spoken corpus show that in Dolgan there is greater acceptance than in Sakha of word orders that differ from this standard constellation, in particular an acceptance of SVO. Importantly, the post-verbal object does not necessarily encode an already activated topic, defocused constituent, or afterthought, as was described for other Turkic languages, but can also occur in pragmatically neutral utterances. To substantiate this statement, in Dolgan 41 out of 175 sentences with overtly expressed verb and object are VO. These 23.4% contrast sharply with the 1.7% of VO -sentences just mentioned for Sakha.

Examples 7.12 and 7.13 are clear instances of objects in the position of an afterthought. In 7.12 the postposed *iñe-ŋ haŋa-tin* ‘your mother’s word’ is the object of *iste-gin* ‘you listen’, and in 7.13 *ol tiri:-gin* ‘that skin’ is the object of *ij-ï:l-lar* ‘they hang’. This interpretation is corroborated by the fact that in both cases the verb is followed by the particle *buo*, which occurs at the end of a clause and is always followed by a pause. It has some kind of assertive meaning, displays a drop in intonation and turns the preceding clause into a closed unit. Everything following this particle is a new sentence, or an afterthought.

DOLGAN

- (7.12) *iste-gin buo, iñe-ŋ haŋa-tin*
 listen.SIM.CV-PRED.2SG PRT mother-POSS.2SG language -ACC.3SG
 ‘You listen to your mother’s word.’ (ESB: 42)

- (7.13) *iti... kimieçe taŋas ija:n-ar kim-ner-ge*
 this... who.DAT clothes be.hung-PRS.PTC who-PRED.3PL-DAT
ij-ï:l-lar buo ol tiri:-gin, taŋas
 hang-PRS.PTC-PRED.3PL PRT that skin-ACC.2SG clothes
ija:n-ar.
 be.hung -PRS.PTC
 ‘Ehm... they hang it on a clothes hanger, the skin, a clothes hanger.’ (ESB: 34)

On the other hand there are examples like 7.14 and 7.15, in which there is no indication that the object is separated from the verb in any sense, even though the object occurs in clause final position. In 7.14, the post-verbal object occurs in the combination *kötöxtö ginini* ‘he lifted him’, which is a clear syntactic and intonational unit. An interpretation of *ginini* as afterthought seems, in the absence of any semantic, syntactic or intonational cues, very unnatural. Rather, 7.14 is a semantically and pragmatically unmarked description of this lifting event, and the changed word order does not affect the interpretation, i.e. this sentence would have exactly the same reading as in a sentence where the order is *ginini kötöxtö*. The same holds for 7.15, where *ontugun* ‘that’ follows the verb *tutuoxta:χχin* ‘you should hold’.

DOLGAN

- (7.14) *hinnan-an* *χann-an* *bar-an* *ke* *de* *ol*
 relax-SQ.CV and.so.on-SQ.CV go-SQ.CV CONTR PRT PRT
kötöχ-tö ***gini-ni,*** *kötöχ-tüler* *kriltso-χa:m-mit*
 lift-PST.3SG 3.SG -ACC lift -PST.3PL doorstep-DIM-1PL
ürdük-ke:n *e-te* *ürdük*
 high -DIM be-PST.3SG high
 ‘After relaxing and so on, well he lifted him, ehm they lifted him, our porch
 was high, high.’ (TJP: 85)

- (7.15) *üčügej-dik* *tut-uoχ-ta:χ-χin* ***on-tu-gun,***
 good-ADVLZR hold-FUT.PTC-PROP-PRED.2SG that -DER -ACC.2SG
ï:p-pat *kördük*
 send-PRS.PTC.NEG similar
 ‘You should hold that well, in such a way that it doesn’t drop.’ (ESB: 74)

Now, how meaningful is this observed difference? Is it only anecdotal, reflecting chance variation, or is the higher frequency of VO-clauses in Dolgan significantly different from Sakha? A quantitative analysis presented in the next section is intended to solve these questions. Three specific questions are addressed: a) Do word order patterns differ significantly across Dolgan and Sakha? b) Do word order patterns differ across text genres? c) Does the difference between Dolgan and Sakha hold for the language as a whole, or is the difference due to idiosyncratic language use of particular speakers?

7.5 QUANTITATIVE COMPARISON OF WORD ORDER PATTERNS IN DOLGAN AND SAKHA

7.5.1 METHODOLOGY

In order to make a quantitative comparison for SOV and SVO across the two languages, I coded transitive sentences for S, O and V in a randomly selected part of the corpora for both Dolgan and Sakha. For Dolgan, 5 narrative texts and 6 Pear Stories were coded for S, V and O, which yielded 512 utterances, produced by 11 different Dolgan speakers. These 512 utterances included intransitive sentences, transitive sentences with an unexpressed direct object and transitive sentences with an overtly expressed direct object. Since only the last category is relevant for the current analysis of V and O order, only these transitive sentences were included, totaling 175 sentences. Transitive clauses for which O was not overtly expressed (as exemplified in 7.16, where the verb has a subject, *iti* ‘this’ and an indirect object *uol oyoχo:nugar* ‘to the little boy’, but no direct object), were excluded from the analysis.

DOLGAN

- (7.16) *iti uol oyo-χo:n-ugar... kim... ber-s-i-bit*
 this boy child-DIM-DAT.3SG who give-RECP-PST.PTC
 ‘To the little boy he... what-is-it-called gave (it).’ (TIS: 11)

For Sakha the total number of coded utterances was 575, taken from 2 long narratives and 6 Pear Stories, narrated by 7 different speakers (1 of whom produced a narrative as well as a Pear Story). 176 utterances were transitive clauses with an overtly expressed object and were included in the analysis. The transitive clauses were further classified as OV and VO order, the frequencies of which were then calculated and compared across the two languages. After that, the potential significance of the frequency difference between Dolgan and Sakha was evaluated with the help of statistical models, which will be described below.

Before discussing the comparison in detail, a few points need to be made. First, so far word order has been discussed in terms of S, O and V. However, as mentioned before, in spontaneous speech the overt expression of S, O and V is the exception rather than the rule, in particular in pro-drop languages like Dolgan and Sakha. Therefore a more general, but more useful global categorisation was made of OV and VO order, where OV includes the theoretically possible patterns OV,

SOV, OSV, OVS, and VO includes VO, SVO, VSO and VOS. Second, since text genre potentially influences the frequency of transitive clauses in general, the text corpus used for this analysis consisted of spontaneous narratives, as well as the semi-spontaneous Pear Stories to control for text genre (see Section 1.2.2 for details). The idea behind this is that a story on the preparation of reindeer hide, or the construction of a boat will naturally include more transitive clauses than a story about one's family, since procedural texts typically involve agents acting on patients, which is the argument scheme for the prototypical transitive clause. Thus, the frequency of transitive clauses may be dependent on the chosen topic of the narrative. In addition, certain text genres may correlate with particular kinds of pragmatic structures, and may therefore favour a more frequent use of certain clause orders. By using semi-spontaneous texts it is possible to control the choice of topic and discourse pragmatic function of the narrative to some extent, thus increasing the degree of comparability across speakers. Although the interpretation of the film used to elicit the Pear Story can of course not be controlled and a certain level of variability will naturally remain, the uniform input considerably and sufficiently limits the divergence of the output (see Section 1.2.2).

Differences in word order frequency can be evaluated in several ways. The most straightforward way would be to compare percentages of occurring orders across languages and across genres, but the downside of this method is that it does not provide any information with respect to the significance of the different percentages. A much better result with regard to this issue can be achieved by applying statistical models, which are designed for dealing with just this task. The best model for the evaluation of the linguistic data in this study is a so-called Generalised Linear Mixed Model (GLMM)¹, more specifically a Poisson model². A GLMM allows you to control for speakers' behaviour by including individual

¹ I am grateful to Roger Mundry, Matthias Urban and Michael Danneman for choosing the correct statistical models and applying them to my data.

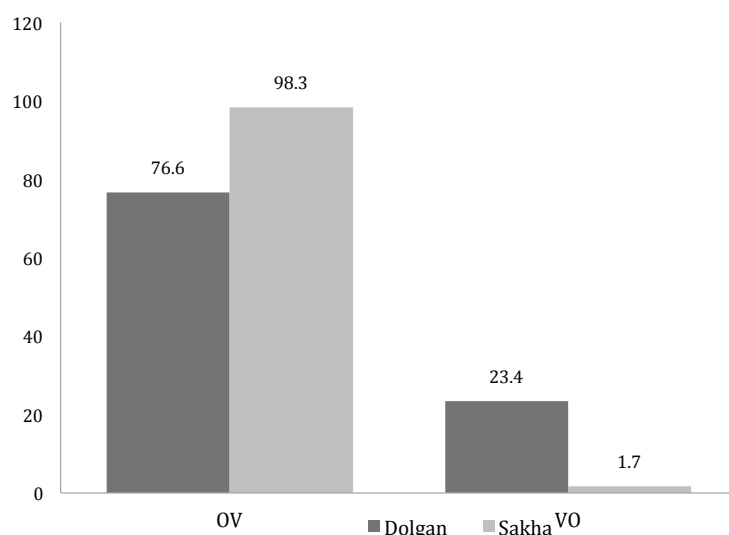
² The Poisson model employed for these calculations used a log-link function and was built in R (R Development Core Team 2009), using code by Bates and Maechler (2010). This means that the data needed to be log transformed in order to fit the model. In this model, the total number of utterances was included as a log transformed offset term, controlling for effort. In a Poisson distribution, the mean is equal to its variance. If the variance is greater than the mean, or if it is dependent on the observed value, we speak of overdispersion and the model would not be appropriate for use. However, in the present study there was no issue with overdispersion ($\chi^2 = 12.94$, $df = 16$, $p = 0.68$, dispersion parameter = 0.81).

speakers as so-called random effects, which are factors that cannot be controlled by the experimenter. The inclusion of such random effects was a necessary requirement to the model, because not all data points were independent of each other, due to the fact that some of them were produced by one and the same speaker. If the statistical model does not account for this fact, there is a danger that the data might be biased in one way or another by the behaviour of an individual speaker. Since the independence of data points is mandatory for standardised models for significance such as chi-square or Fischer exact tests, these could not be used for the current purpose. Other factors that were implemented into the model as fixed effects were language (Dolgan vs. Sakha) and genre (Pear Story vs. narratives).

However, for a legitimate application of a GLMM to these data, it was necessary to carry out a test for model stability first. In principle, GLMMs also work only with independent data points. Since the present data set contained one non-independent data point (i.e. the speaker who participated in the production of language data for both genres) it needed to be proven prior to the choice for the Poisson model that the effect of this data point was no different from the effect of the independent data points. For this purpose a Generalised Mixed Model was used, showing that the single non-independent data point did not cause any different effects than the independent ones. Therefore, the GLMM model could be applied with clear conscience.

7.5.2 RESULTS

Figure 7.1 shows a summary of the distribution of OV and VO order across Dolgan and Sakha. In this figure, data from spontaneous narratives and Pear Stories were collapsed because it turned out that text genre had no significant effect on the clause order in Dolgan or Sakha (see below for details). The proportions of OV and VO order are calculated relative to the total number of transitive clauses with overt expression of V and O in the texts. The dark grey columns reflect the proportions in Dolgan, whereas the light shade represents the results for Sakha.

Figure 7.1 Proportions of OV and VO-order in Dolgan and Sakha

As can be seen from this figure, there is a noticeable difference in proportions of OV and VO occurrence across the two languages. In Sakha, the ratio of OV, the typical word order for the Turkic language family, is 98.3%, whereas for Dolgan this is only 76.6%. On the other hand, VO order occurs only in 1.7% of the clauses in Sakha, whereas the proportion of 23.4% in Dolgan is much higher. The exact numbers, specified for language as well as for text genre (i.e. for spontaneous narratives and Pear Stories separately) are provided in Table 7.1 below.

Table 7.1 Numbers and proportions of OV and VO clauses in Dolgan and Sakha

Word order	Dolgan				Sakha			
	OV		VO		OV		VO	
	no.	%	no.	%	no.	%	no.	%
Narrative	93	78.2	26	21.8	101	98.1	2	1.9
Pear Story	41	73.2	15	26.8	72	98.6	1	1.4
Total	134	76.6	41	23.4	173	98.3	3	1.7

The numbers in the table suggest that the text genre does not have much influence on the distribution of OV versus VO orders. In Dolgan, OV order is attested for 78.2% of the transitive clauses in spontaneous narratives, and 73.2% in the semi-spontaneous Pear Stories. For VO order a comparable similarity between

the text genres is found, namely 21.8% for narratives, and 26.8% for the Pear Stories. In Sakha, OV order occurs in 98.1% of all overt transitive clauses in spontaneous narratives, and 98.6% in the Pear Stories. VO clause order occurs in 1.9% in the narratives, and 1.4% in the Pear Stories. While these numbers suggest homogeneity in word order distribution across text genres, and a different distribution across languages, statistical tests are needed to establish whether this intuition is correct, in other words, whether the frequency differences between languages and across text genres are significant or whether they are likely to reflect chance variation.

The calculations made by the GLMM Poisson model reveal that the frequency of occurrence of VO order indeed differs significantly across Dolgan and Sakha. It shows that VO order occurs significantly less in Sakha than in Dolgan. This is evidenced by the significance of the so-called estimate value, which is an estimation made by the model with respect to the relative frequency of a certain result in Sakha and Dolgan (estimate = -2.63, $p < 0.0001$). At the same time they confirm that genre makes no significant difference for the occurrence of OV or VO order within languages ($p = 0.599$) suggesting that the proportions of OV and VO are stable, regardless of whether the text was a spontaneous narrative or a semi-elicited Pear Story.

Finally, a stability test was carried out to see whether one of the data points, i.e. speakers, could be the cause of the observed patterns. This is tested by constructing models in which one of the data points is removed at a time, and comparing the range of estimates for those models with that of the original GLMM. This shows neither a strong impact on the estimate for the differences between languages (range from -3.11 to -2.47, estimate for the original model: -2.63) nor on that for genre (range from 0.01 to 0.33; estimate for the original model: 0.165).

Summarising we can conclude that Dolgan and Sakha differ significantly from each other with respect to the frequency of VO clause order, regardless of speaker and of text genre. The proportion of atypical VO clauses is much higher in Dolgan than in Sakha, which could foreshadow an ongoing change in Dolgan. The next section will be concerned with possible explanations for this development.

7.6. DISCUSSION

Three possible scenarios come to mind for the explanation of increased VO order in Dolgan. First, it could be a language-internal development. Second, it could be motivated by contact with Russian, and third, it could have developed under the influence of contact with neighbouring indigenous languages. Of these three possibilities, the third can quickly be dismissed, since the neighbouring Tungusic language (Evenki), as well as the Samoyedic languages (Nganasan, Enets, Nenets) have the same SOV basic word order as Dolgan. Therefore, influence of these languages in this linguistic domain would have no noticeable effect. As a result, a stimulating role of Tungusic or Samoyedic speakers with respect to the increase of SVO order in Dolgan can be confidently excluded.

7.6.1.1 LANGUAGE-INTERNAL MOTIVATIONS FOR WORD ORDER CHANGE

With respect to language-internal development, a change from SOV to SVO is cross-linguistically not uncommon. Dik (1997) explains this in terms of the ‘Principle of Increasing Complexity’. In his words, this means that there is a preference for ordering constituents in an order of increasing complexity’ (Dik 1997: 404), where the concept of ‘complexity’ roughly corresponds to concepts like the ‘Gesetz der Wachsenden Glieder (‘law of increasing parts’) formulated by Behaghel (Behaghel 1909: 139) or ‘heaviness’ (Hawkins 1983: 90, Mallinson-Blake 1981: 158). In Hawkins’ terms, ‘heaviness’ is a composite notion defined in terms of: a) the length and quantity of morphemes; b) quantity of words; c) syntactic depth of branching nodes; and d) inclusion of dominated constituent. The heavier a constituent, the more likely it is to be placed to the right of the head of the clause. Dik (1997: 410) even dedicates a special ‘principle’ to it, namely ‘Specific Principle 6’, which states that

[t]he Prefield is universally less hospitable to complex material than the Postfield.
Prefield languages may thus be expected to possess strategies for relieving the
Prefield of excessive complexity.

In this quotation, the prefield is the position preceding the head, and the postfield the position following the head of the sentence. He goes on to say that one can

distinguish between 'strict' and 'liberal' prefield languages (that is, SOV languages), where the more liberal prefield languages allow for a certain amount of 'leaking' of constituents beyond the head (Ross 1973, as cited by Dik 1997: 410). This, he argues, may diachronically lead to a gradual change from an SOV language into a 'Prefield-derived SVO language'. These are SVO languages, which have retained a number of prefield properties, such as Karaim, which has adopted the relative pronouns and right branching relative clauses from Russian, but has kept participles and converbs to the left of the verb (Johanson 2002a: 131-137).

Although the preceding account has shown that several scholars employ concepts such as 'heaviness' or 'complexity' to explain tendencies in internally motivated language change, there is no unequivocal explanation for the existence of the principles themselves. For a long time it was assumed that the attested tendency was motivated by general psycholinguistic principles and constraints on processing. Left branching structures in general were assumed to put a heavier burden on memory in production (Yngve 1961, in Johanson 2002a: 120) as well as in comprehension, and would therefore be disfavoured. The argument is that in production the speaker needs to plan the entire sentence before he can even start producing it, due to the fact that the head is in final position. Likewise, the hearer needs to remember all the details and modifications before the eventual head is revealed at the end of the clause. However other studies show that the processing complexity of left-branching structures is not any greater than for right-branching structures (Frazier & Rayner 1988, in Johanson 2002a: 120).

Despite contradictory results in psycholinguistic research, the tendency in languages to position longer and more complex constituents towards the end of the sentence, and thus of 'leaking', remains a fact. Because of this natural tendency in one direction, it is not surprising that the opposite direction of internally motivated language change, from SVO to SOV, is cross-linguistically less common. There are even claims that go as far as to say that a change from SVO to SOV can only occur as a result of contact, and would never happen as a language-internal process (Ross).

These facts leave open for consideration the possibility that the increase in SVO clause order in Dolgan could be a language-internal development, following a universal tendency in language change. However, as can be seen from examples like 7.14 and 7.15, heaviness or complexity can certainly not always be adduced as a motivation for SVO order. The object pronouns are in fact, apart from omission, the shortest possible way of expressing objects. In addition, if it were a change

independent of contact, one would expect it to be pervasive throughout the whole area where the language is spoken, and to occur irrespective of geographical location or sociolinguistic setting. However, the next section will show that this is not the case.

7.6.1.2 LANGUAGE-EXTERNAL MOTIVATIONS APPLIED TO DOLGAN

An investigation of the distribution of VO-order across the villages reveals that its frequency varies depending on geographical location. More specifically, it correlates with the sociolinguistic situation prevailing at the geographical location with respect to the use of Dolgan and Russian. As was described in Section 1.3.2, the villages where I recorded the Dolgan narratives differed considerably with respect to the balance of linguistic dominance between Dolgan and Russian, as well as the attitude towards use of each language. It was mentioned that the Dolgan language is most vital in the villages that are furthest away from the Russian-dominated centers, and that its use gradually decreases as one comes closer to the towns, in particular Dudinka. As can be seen from Table 7.2, the proportion of Turkic OV and Slavic VO-clause order in the speech of the language consultants correlates with this difference in sociolinguistic setting, in particular with the increase of Russian dominance.

Table 7.2: Percentage of OV and VO clause order per community

	OV	VO
Syndassko	90.1%	9.9%
Kheta	70.0%	30.0%
Dudinka	70.7%	29.3%

To put it concretely, there is an increase in VO structures when travelling from east to west, i.e. from Syndassko over Kheta to Dudinka. In Syndassko, VO structures constitute 9.9% of the transitive sentences, in Kheta 30.0%, and in Dudinka 29.3%. These results show that the higher occurrence of VO structures in communities with a strong social and linguistic representation of Russian could be due to transfer of such structures from the dominant Russian into non-dominant Dolgan.

Another possible factor influencing the distribution of word order patterns could be speaker age. If apparent time³ gives a realistic representation of ongoing language change and if one follows the general idea that children are the main locus of language change, or the less widespread idea that innovative structures are predominantly promoted by preadolescents (Ross forthcoming), one would expect younger speakers use innovative SVO structures more frequently than older speakers do. Since it was not possible for me to find enough speakers in each age group for a reliable sample, it is currently not possible to say with certainty whether age plays a role or not. However, the impressionistic data shown below go against this expectation.

Table 7.3: Percentage of OV and VO clauses per age group

Age	Location	OV	VO	Total OV	Total VO
75	Dudinka	70.7%	29.3%	75%	25%
	Syndassko	84.2%	15.8%		
40	Kheta	70.0%	30.0%	77.3%	22.7%
	Syndassko	92.9%	7.1%		
14	Syndassko	93.3%	6.7%	93.3%	6.7%

Table 7.3 includes three age groups, roughly corresponding to three generations, and the percentage of OV and VO clauses they produced in their spontaneous speech. Due to the labour-intensiveness of manual word order counting, the number of individuals per age group is only two. For the youngest age group only one individual was included because the narratives from the other children in this age group were unsuitable for the current purpose, either due to the absence of transitive clauses, or to interference of the parent. The table shows that the distribution of OV and VO order within the age groups is far from homogenous. In the age group of 40, one speaker uses VO order in 30.0% of the transitive clauses, whereas the other uses it in only 7.1%. A similar situation, though less extreme, applies to the age group of 75. Since the number of individuals is so low, this diversity could of course be due to chance, but it seems that word order patterns cannot be correlated with a particular generation. Second, the average frequencies

³ 'Apparent time' is the idea that language variation between speakers of different age groups at a particular moment in time is representative of the development of the language through time. According to this idea, synchronic language variation can be used to study diachronic language change (Labov 1994: 28-29, Chapter 3).

for age groups (found in the columns 'total OV' and 'total VO') are not distributed across these groups in the way expected for ongoing language change, if credibility is given to apparent-time predictions⁴. This could be taken as an argument against language-internal innovation for this particular feature in Dolgan.

7.6.2.1 LANGUAGE CONTACT AND WORD ORDER CHANGE

As it turns out, the literature about the nature of word order change is ambiguous. From the perspective of language contact studies, word order is characterised as a linguistic feature that is affected in contact situations relatively easily, whereas the literature on language acquisition classifies word order as a 'deep' structural feature that is supposedly very resistant to influence from other languages. Thomason and Kaufman write about word order change in contact situations:

The evidence we have collected does not support the often implicit assumption, in the literature on word order change, that word order patterns constitute a fundamental deep structural feature relatively impervious to foreign influence. On the contrary, word order seems to be the easiest sort of syntactic feature to borrow, or to acquire via language shift. (Thomason and Kaufman 1988: 54-55)

They explain this by the idea that SOV and SVO patterns both fulfill the syntactic function of identifying a subject and object with respect to each other and to the verb. They illustrate their argumentation with examples from Finnish, which changed from SOV to SVO order under the influence of Indo-European languages and Austronesian languages of New Guinea, which show a change in the opposite direction (SVO to SOV) due to contact with Papuan languages. These cases seem to be instances of heavy copying of structure in a situation of language maintenance, and as far as I can tell they do not give examples of word order change in language shift situations. However, the message remains valid nonetheless: word order patterns do change under the influence of contact.

⁴ I realise, of course, that a larger sample of texts and informants would potentially provide a different picture. Multiple speakers for every location would have been desirable, but were not always possible to find.

This view is confirmed in later work by Thomason (2001: 88) in which she repeats that basic word order is among the most common features that are affected by structural interference. While she labels this phenomenon primarily as a “replacement of native linguistic features by new interference features” (ibid.: 87), she adds the possibility that change in word order may be the result of convergence, which she defines as “any process through which two or more languages in contact become more like each other” (ibid.: 89). The term convergence for her implies that it is impossible to clearly define a source language and a recipient language. Rather, the languages converge towards each other, being both source and recipient language at the same time. In this context she adduces the example of Kadiwéu, spoken in Brazil, which in natural discourse shows six different word orders (including SVO), but in translations from Portuguese displays an unusually high frequency of SVO word order, copying the unmarked Portuguese order of constituents (Sandalo 1995 in Thomason 2001: 89). Although the adaptation is unidirectional, Thomason prefers to characterise this case as convergence, since SVO word order was already present in Kadiwéu, and therefore it would be inappropriate to call Portuguese the source language and Kadiwéu the recipient language. According to Thomason, this does not necessarily represent a change in Kadiwéu, although she admits that it could eventuate in it, but rather is an example of how changes can start through a shift in frequency of particular constructions⁵ (see also frequential copying (Johanson 2002b: 292, Heine and Kuteva 2005: 47) as discussed in Section 6.4).

This is in line with the ideas formulated by Thomason (2001: 69), who describes word order as among the “next easiest things to borrow”, after the lexicon. However, Heine (2008) argues that in fact there is no case in which ‘new’ word order is completely new and unprecedented in the language:

What frequently happens is that speakers draw on a minor use pattern – one that has a more marginal status, being used rarely and/or only in specific contexts only to build a new major use pattern by increasing the frequency of use and extending the range of contexts in which it may occur. (Heine 2008: 55)

⁵ While Kadiwéu in this process is indeed (its own) source and recipient language at the same time, enhanced by contact with Portuguese, the change still seems to take a one-way direction from Portuguese to Kadiwéu. Whether Kadiwéu influences Portuguese in other linguistic domains is not further specified.

He admits that often this process does not lead to a complete change in word order, in which case it may result in a more flexible word order in the recipient language. As an example he gives Eskimo speakers of North America who are in contact with English speakers. He observes that instead of changing their own word order pattern completely on the strict SVO model of English, Eskimo word order has simply become more free (Heine 2008: 57).

These accounts give the impression that word order change is rather common and easily achieved in language contact situations, and the considerable number of case studies provides supporting evidence. However, these accounts are relatively unspecific with regard to one or more of the following factors: the sociolinguistic conditions in which they occurred (language maintenance or shift), the identity of the initiators of change (L1 or L2 speakers), and the underlying processes of change (borrowing or imposition). As was pointed out in Chapter 3, the combination of these factors is important for an accurate description and analysis of any contact situation and its linguistic outcomes (see Sections 3.1.3, 3.1.4). Through the complex interplay of each of these factors, different contact situations may lead to the same surface outcomes, and reversely, comparable contact situations may lead to different outcomes. Only detailed sociolinguistic information will make it possible to describe and/or reconstruct the events that underpin the outcomes of language change in the most realistic way.

7.6.2.2 LANGUAGE ACQUISITION AND WORD ORDER CHANGE

The literature on the acquisition of syntax presents a different view of word order change. Part of this may be caused by the fact that much of the work in this field has been dominated by ideas from generative linguistics. As was briefly mentioned before, a common assumption in this research tradition is that word order belongs to the so-called ‘deep structure’ of language, and thus cannot be changed after the relevant ‘parameter’ for this feature has been set (e.g. Lightfoot 1979). One argument in favour of this assumption comes from studies on L1 attrition in individuals who have lived abroad for a long time and hardly ever use their L1. Several studies show (Schmid 2002, Altenberg 1991 in Lucas 2012: 282) that the difficulties these individuals experience in comprehension and production of their L1, including word order, may be only “the result of temporary difficulties with access and retrieval”. (Lucas 2012: 281). According to Lucas this indicates that in

fact the parameter setting, or competence, of the bilingual speaker has not changed, but only that the performance has changed temporarily due to high activation of L2 and low activation level of L1. Since in this view word order is strictly tied to parameter settings that are fixed in early childhood, the implication is that word order change can only be initiated by young infants, at a time when this parameter is set differently from that in the language system of their parents. The consequence of this idea would be that contact must be excluded as a potential cause of word order variation and change, because the only source of change (the infant) does not exactly actively engage in influential social contact, and it does not have the necessary network ties for new language variants to be spread across the community. Although Lightfoot does recognise language variation and contact as possible external factors in the process of language change (Lightfoot 1979: 374), the pathway by which he proposes that contact-induced variants reach, and eventually settle into the grammar of the infant is not entirely clear.

Additional work on language acquisition and change supports the skepticism towards the monopoly of infants in language change. For example Aitchison (1981: 180) supports the idea mentioned above that babies cannot be the main source of language change since they do not have the social significance nor the network ties needed for new variants to become favourable over others and spread through the community. Instead, she argues that language variation is only meaningful for the field of language contact when it occurs in children from 4 years onwards, when they begin to engage in social activities, and identify with certain social groups, which may be different from their parents. She puts special emphasis on preadolescents and adolescents because these are the groups whose members are easily influenced by peers and people a little older than themselves, but are, counter to generative convictions, still able to make fundamental changes to their language (c.f. Light Warlpiri and Gurinji Creole in contact with English based Aboriginal Kriol (O'Shannessy 2005, McConvell and Meakins 2005), German in contact with English (Clyne 1992), all examples adduced by Ross forthcoming).

As will be recalled from Section 3.1.5.2, this view is strongly supported by Kerswill's work, which shows that language changes throughout a person's lifespan, but that the different kinds of change are conditioned by a person's life stage. Partly in line with Aitchinson's conclusions, Kerswill attributes the greatest significance in the emergence of innovative grammatical patterns to preadolescents, and not to adults or in infants (Kerswill 1996: 198). While Kerswill's

studies focus on the emergence of differences between dialects, in a recent paper Ross applies similar ideas in his explanation of change in contact situations involving different languages, when he argues that preadolescents are also crucial agents in the initiation of contact-induced grammatical change, including calquing of word order patterns (Ross forthcoming).

To recapitulate, the dominant idea in the acquisition literature that word order belongs to ‘deep structure’ and therefore cannot be changed after infancy, thus tacitly implying that word order cannot change due to contact, is contradicted by multiple case studies. This, in combination with evidence from language contact theory now opens the way for an account in terms of language contact to explain the word order variation in Dolgan.

7.6.3 WORD ORDER VARIATION IN DOLGAN EXPLAINED

Returning to the data from Dolgan, it is clear that it would be wrong to assume that the increase in SVO structures is due to language-internal factors alone. The observations that the heaviness principle does not always apply in SVO structures and that SVO order does not correlate with age in the expected direction, in combination with the fact that high SVO frequency is found in an area of intense contact with an SVO language, argues against this explanation, and in favour of an account in terms of contact. Of course, the observed tendency to develop SVO structures through language-internal change may certainly have enhanced this process in Dolgan, but taking the fact that this change is cross-linguistically common as the single explanation, would ignore an obvious and significant aspect of the story, which became clear from the correlation shown in Table 7.2.

The fact that SVO order was already an available, but pragmatically marked, structure even before contact with the Russians intensified also facilitated the extension of this construction into less marked contexts (see Johanson 2002a: 111-112, 2002b: 292, Heine 2008: 31, 43, 56-57). Nonetheless, contact with Russian seems to have been the main trigger for the introduction of the option of VO word order. Accepting contact with speakers of Russian as a primary explanation for this difference between Dolgan and Sakha, questions arise with respect to a) the relative status of the languages in contact; b) the initiators of the change (children, adults or preadolescents, L1 or L2 speakers of Russian?) and c) the process underlying the change (borrowing, attrition, imposition?).

The sociolinguistic situation on the Taimyr Peninsula leads me to think that this development in Dolgan must be the result of more than just one process of change on the level of the bilingual individual. With respect to the initiators of change, it was not possible to identify one age group in which SVO occurs consistently more than in others (see Table 7.3). Rather the innovative word order patterns seem to occur in speakers of all age groups who live in a Russian-dominated environment, and for whom Russian has become their dominant (in Van Coetsem's terms), and most highly activated (in Lucas' terms), language. In the remaining part of this discussion, I will focus on this group of speakers only.

While the individuals in this category are all Dolgan people who are in a Russian-dominant environment, even this group is anything but homogeneous, and includes people with very different levels of proficiency in Dolgan. Following Van Coetsem's theory, this would mean that the same result (SVO sentence structures) can be explained by two different processes of change depending on the linguistic dominance (typically correlating with age) of the speakers. The argumentation for this is rendered schematically in Table 7.4.

Table 7.4: Linguistic dominance and processes of change in different age groups

	Age group	
	> 70	< 40
L1 (dominant)	Dolgan	Russian
L2 (non-dominant)	Russian	Dolgan
Direction of transfer	L2 → L1	L1 → L2
Agentivity	L1	L1
Process of change	Borrowing	Imposition

Of course such a differentiation can only be made on the level of the individual, and while the change is in progress. Although it was argued before that age does not play a role in the *frequency* of occurrence of SVO structures, it is indirectly a distinguishing factor when it comes to the *process* underlying the appearance of these structures in Dolgan, because of its link with linguistic dominance. In the table above, the youngest age group, including children and teenagers, has been left out. Their Dolgan did not display much influence from Russian in the most isolated village of Syndassko, and in other villages children do not speak Dolgan anymore, thus making influence from Russian complete and predicting a shift to Russian in these communities in the near future.

The first group consists of the oldest generation (average age 75), who grew up in the 1940-50's. Although Russian influence was already strong on the Taimyr in this period (e.g. children were often forbidden to speak their native language in public), children would still have been brought up by Dolgan-speaking parents, who were often monolingual in this language. Thus one may assume that their L1 and dominant language was Dolgan in infancy and early childhood, and that they later learned Russian as an L2 in school. Despite the increasing presence of Russian in their community, the Dolgan language would remain for them an important means of communication in the interaction with their parents and other members of the community in non-public settings. Disregarding individual exceptions for the sake of generalisation, one can say that this generation is bilingual in Russian, but has remained dominant in Dolgan, regardless of occasional higher activation levels of Russian, which are situationally conditioned. Against this background, the presence of SVO word order in their variety of Dolgan can be best explained through the process of borrowing, more specifically structural borrowing occurring in a situation of intense contact (Thomason and Kaufman 1988 and Section 3.4.1 of this thesis). Structures from the non-dominant source language (Russian) are transferred to the dominant recipient language (Dolgan) due to high exposure to the source language in the community. Possibly this happens to reduce processing costs for the speaker, as well as for the hearer for whom Russian is most probably also the most accessible language.

A different situation holds for the younger speakers (40 and younger). Growing up in the 1970's and later, these Dolgan individuals had bilingual parents, and were mostly settled in Russian-oriented villages. Even if they spent the first few years of their life in the tundra, they were brought to boarding school from the age of 5 where any initial Dolgan dominance would quickly disappear. The boarding schools were monolingually Russian, and the use of indigenous languages was not at all appreciated, if not forbidden. This led automatically to a change in attitude towards both languages. Russian was represented as prestigious and the language of education and development, and children would speak it to their teachers, but often also to each other. In the beginning they did this perhaps mainly so they would not get 'caught' speaking an indigenous language by a teacher, but later possibly because Russian became the more activated and therefore easier language, so that in addition to the social issues, retrieving Dolgan would mean greater psycholinguistic effort. For these people it is highly questionable whether Dolgan can be called their L1 and it is certainly not their

dominant language. They speak Russian like native speakers, whereas their use and knowledge of Dolgan is more limited and not as fluent. If they do speak Dolgan, code-switching with Russian is common. Therefore, for this generation Dolgan has often acquired the status of a non-dominant L2. It goes without saying that also to these statements there are exceptions, but the purpose here is to characterise the general tendency.

I deliberately chose to talk about L1, L2 and dominant language instead of referring to Russian or Dolgan as the 'native' language for the following reasons. Despite the often poor knowledge of Dolgan in the younger generation, many Dolgans would still say that their native language is Dolgan. This is completely justified considering the fact that they are Dolgans, and it was one of the languages they grew up with from birth. However, this choice seems to be based rather on factors such as ethnic identity and association with a certain ethnolinguistic group than on actual linguistic proficiency. The sociolinguistic features of the group under forty leads me to the conclusion, that word order change in these people (i.e. people whose dominant L1 is Russian and whose non-dominant language is Dolgan) is the result of the process of imposition. Their high exposure to, and psycholinguistic dominance in Russian causes them to project sentence structures from their L1 (Russian) onto their non-dominant L2 (Dolgan).

The differentiation between borrowing and imposition can only be made at the level of the individual speaker, while the change from SOV to SVO is in progress. Once it is completed, the detailed information on individual variation will no longer be available. Therefore, diachronically, and at the community level, this change is best explained by imposition by Dolgans who are dominant speakers of Russian.

8.1 INTRODUCTION

8.1.1 DEFINING COORDINATION AND SUBORDINATION

Clause combining (Haiman & Thompson 1988) deals with the relation between clauses (Fabricius-Hansen & Ramm 2008: 2). It is also known as clause linkage (Lehmann 1988: 181), whereby a clause is defined as a finite or non-finite verb phrase (Fabricius-Hansen & Ramm 2008: 6) or “any syntagm containing one predication” (Lehmann 1988: 182). Traditionally, the types of relations between the clauses are further subdivided into coordination and subordination (Cristofaro 2003: 15). Usually this distinction is made mainly on the basis of formal criteria of morphosyntactic (a)symmetry. This focus on the morphosyntactic component can be attributed to the fact that the study of clause combining was dominated for a long time by generative linguistics, a theoretical framework in which semantics and pragmatics played only a marginal role. In a purely formal sense, coordination would then be defined as a symmetrical construction “in which all of the constituents are of the same syntactic category and this is also the category of the whole construction” (Haspelmath 2004: 33). Subordination on the other hand would be an asymmetrical construction “in which the category of the whole construction is determined only by one of the constituents (the head), while the

other constituents (the dependents) play no role in this respect.” (ibid.) Examples of both constructions are given in 8.1 and 8.2 respectively (Comrie 2008: 3).

(8.1) [John plays the flute] and [Mary sings madrigals]

(8.2) Columbus thought [that the earth was round]

However, cross-linguistic investigation as well as more in-depth study of well-known languages such as English, have shown that formal syntactic criteria alone are not sufficient to account for the typological diversity of clause combining constructions cross-linguistically. Moreover, Comrie argues (among others) that constructions cannot always unambiguously be classified as being clearly coordinate or subordinate. To give an example from a familiar language, Comrie (2008: 3) points out that English has structures which are syntactically coordinate, and yet show certain features of subordination. Similarly, there are structures which are syntactically subordinate but behave as if they were coordinate. Finally, he discusses the situation where one and the same syntactic construction can be classified as coordinate as well as subordinate depending on interpretation.

The first case is illustrated by example (8.4a), which seems to have the same syntactic structure as (8.3a) but behaves differently with respect to one typical feature of coordinate constructions, namely the Coordinate Structure Constraint. This constraint prevents constituents of a single conjunct of a coordinate construction to be relativised (Comrie 2008: 3). Despite the fact that 8.3a and 8.4a have the same syntactic structure, relativisation of one of the conjuncts in 8.3b is ungrammatical, whereas this is acceptable in example 8.4b (Comrie 2008: 3-4).

- (8.3) a. [John plays the flute] and [Mary sings madrigals]
 b. *The madrigals [that [John plays the flute] and [Mary sings-]]

- (8.4) a. I went to the store and bought a book
 b. The book [that I went to the store and bought-]

The second issue is illustrated by the grammaticality of tag questions in English. Normally tag questions are only applicable to main clauses and not to subordinate clauses (8.5). However, examples like 8.6c show that there are exceptions to this rule.

- (8.5) a. Columbus thought that the earth was round
 b. Columbus thought [that the Earth was round], didn't he?
 c. *Columbus thought [that the Earth was round, wasn't it?]
- (8.6) a. I think that John is leaving tomorrow
 b. ?I think [that John is leaving tomorrow], don't I?
 c. I think [that John is leaving tomorrow, isn't he?]

As in the previous example, these sentences have the same formal syntactic structure, yet behave differently with respect to this particular syntactic operation. Tagging the subordinate clause in 8.5c is ungrammatical, as predicted, but in 8.6c the same operation yields a grammatical result, and more than that, it is better than tagging the syntactic main clause 'I think', which is pragmatically odd¹. An illustration of the third problem is provided by juxtaposition, which is a very common way to link clauses in many languages (Sampson, Gil and Trudgill 2009). This strategy is syntactically symmetrical but can often be interpreted as coordinate as well as subordinate, depending on the context. Comrie illustrates this phenomenon with examples from Haruai, but, as will be discussed below, it is a common clause linkage strategy in the languages of Siberia as well, including Dolgan, Sakha and Evenki. Example 8.7 shows two syntactically independent clauses, the relation between which is typically interpreted as temporal or as conditional due to their position and semantic content.

HARUAI

- (8.7) *Rwö* *watk* *h-ön-a*, *an* *hölm-n-ŋ-a*
 environment evening come-FUT.3SG-DECL we sleep-FUT-1PL-DECL
 'When the evening falls we will sleep.' (Comrie 2008: 13)

Thus, the examples above show that there can be a mismatch between the syntactic structure of a construction and its possible semantic or pragmatic interpretations (Yuasa & Sadock (2002) in Haspelmath 2004). The fact that such a

¹ Diessel and Tomasello argue that mental verbs in English (such as I think, I bet, etc.) may have a special status, since they are often used as a formula. Instead of viewing such sentences as consisting of two propositions, they argue that the syntactically embedded clause should be analysed as the main clause, and that the mental verb serves as a modifier, in a way comparable to an adverb (e.g. apparently). See for more discussion of this issue Limber (1973), Thompson (2002), Diessel and Tomasello (2001), Stapert (2009).

mismatch is possible implies that subordination and coordination should not be defined purely on the basis of their syntactic properties, but that semantics and pragmatics play an equally important role. The recognition of this fact has led to a reconsideration of the notions of coordination and subordination, whereby both linkage types are assigned a syntactic as well as a semantic component, the status of which may, but need not coincide. Haspelmath defines the difference between coordination and subordination as:

A construction [A B] is considered coordinate if the two parts A and B have the same status [...], whereas it is not coordinate if it is asymmetrical and one of the parts is clearly more salient or important, while the other part is in some sense subordinate. (Haspelmath 2004: 3)

While Haspelmath recognises that this definition is rather general and needs further specification, it allows for a syntactic as well as for a semantic or cognitive interpretation of (a)symmetry, and is in many cases able to solve mismatches such as the ones described above. In addition, it accounts for cases of clause-internal coordination, such as so-called pseudo-coordinate constructions as in 8.8, where the morphosyntactic structure of the conjuncts is subordinate, but their semantics are coordinate, as can be seen by the plural agreement on the verb (for more detail see Section 8.2.3.3).

RUSSIAN

- (8.8) *Saša s Mašej pošli v kino*
 Sasha with Masha.INST go.PST.PL in movie
 “Sasha and Masha went to the cinema.”

The interaction between clause linkage types and syntactic or semantic (a)symmetry is summarised in Table 8.1. As can be seen from this table, it is the semantic and not the syntactic (a)symmetry that correlates with the distinction between coordination and subordination, and thus the semantic or cognitive factors can be said to overrule the syntactic factors (such as embedding) in case of a discrepancy between the two (e.g. pseudo-coordination).

Table 8.1 Interaction between syntactic and semantic factors in the determination of clause linkage

	type	
	Syntactic symmetry	Semantic symmetry
Coordination	+	+
Pseudo-coordination	-	+
Subordination	+	-
	-	-

Thus, a broader definition that includes syntactic as well as semantic factors accounts better for the cross-linguistic diversity of clause combining constructions than a definition based on syntactic criteria alone.

Nonetheless, even this definition leaves certain structures unaccounted for, and certain voices in the literature have expressed serious doubt as to whether a strict dichotomy between coordination and subordination can be upheld at all (Lehmann 1988, Cristofaro 2003, Haspelmath 2004, Comrie 2008). Instead of supporting the bipolar system in which a construction either belongs to the coordinate or subordinate category, Lehmann proposes a clause linkage continuum, where coordination and subordination form the extremes of a gradient scale. Some constructions are classified as unambiguously coordinate or subordinate (the ones for which the syntactic and the semantic component point in the same direction), other constructions can be placed at any position on the continuum, some closer to the coordinate extreme, others closer to the subordinate extreme, depending on the syntactic and semantic properties of that particular construction. Along similar lines, Comrie states that “the opposition [of coordination and subordination, E.S] is a question of degree rather than a strict dichotomy” (Comrie 2008: 16). Haspelmath concludes his discussion of the matter with:

It remains difficult to operationalize the basic undisputed intuition that coordination involves symmetry, while subordination involves asymmetry. There are many constructions showing mixtures of both, and we are only at the beginning of understanding what constraints there might be on such mixtures. (Haspelmath 2004: 37)

While the theoretical issues concerning coordination and subordination clearly have not been solved and the search for a characterisation that fully captures

cross-linguistic variation is still ongoing, the available definitions are sufficient to serve descriptive purposes of individual languages. The following section will give a brief overview of the terms that will be used in the remainder of this chapter.

8.1.2 TERMS AND DEFINITIONS

The definitions of coordination and subordination that I will use in this chapter are taken from Haspelmath (2004) and Cristofaro (2003). Both authors base their distinction on both semantic and syntactic criteria, but acknowledge that the semantic component is decisive in whether a construction is coordinate or subordinate, as was concluded on the basis of Table 8.1.

Coordination is defined as the linkage of two cognitively independent linguistic units, called *coordinands* or *conjuncts*. In principle they can be words, word groups, clauses, state of affairs, but since this chapter is on clause combining, the main focus will be on the unit of the clause. As mentioned above, coordination is characterised by a symmetrical relation between the coordinands and can be expressed *syndetically* or *asyndetically*. In *syndetic* coordination, the relation between conjuncts is established by an overt coordinating element, which is called a *coordinator*. The main types of coordinate relations are a) coordination (roughly corresponding to ‘and’-relations, including enumeration, temporal coordination, specification); b) adversive coordination (‘but’); c) disjunction (‘or’) and d) causality (‘therefore’). In *asyndetic* coordination, the conjuncts are juxtaposed without the presence of an overt coordinator to specify the nature of the relation. Instead, the connection between conjuncts is expressed by means other than morphology, including intonation, the semantic content of the coordinands and discourse pragmatic implications.

Subordination on the other hand is a relation between linguistic units which involves cognitive dependency. In contrast to coordination, subordination is characterised by a relation of functional asymmetry between the two linguistic units (the so-called *Asymmetry Assumption*, Cristofaro 2003: 29) whereby the profile of one of the linked elements is overridden by that of the other. As for coordination, this asymmetrical relation can be established *syndetically* through overt subordinators, as well as *asyndetically* through juxtaposition. Subordinate relations can be subdivided into a) *adverbial subordination*, b) *complement clauses* and c) *relative clauses*.

In this chapter, I investigate clause combining strategies in Dolgan. Since clause combining strategies in Dolgan and Sakha generally coincide, my aim is not to give an exhaustive description of all clause combining strategies in Dolgan, but rather to highlight those aspects of Dolgan clause linkage that differ from Sakha.

To anticipate some of the conclusions, the chapter will show that most of the attested differences are attributable to influence from Russian. In some cases this influence is directly visible through the presence of Russian coordinators in Dolgan discourse, or the flexibility of subordinate clause position with respect to the main clause. In other cases, the variation in Dolgan is argued to be the result of ongoing language attrition, which in turn is the result of a progressive shift to Russian. Therefore, even these changes could be argued to be the result of Russian contact, albeit indirectly. Section 8.2 discusses coordination strategies and Section 8.3 deals with subordination. In each section, an overview is given of the clause combining strategies in Sakha, followed by similar information for Dolgan, after which the differences from Sakha are highlighted and discussed. Section 8.4 pays special attention to the use of Russian coordinators in Dolgan discourse and embeds this phenomenon into existing theories on this matter. The chapter is concluded with a summary and a possible interpretation of the results (8.5).

8.2. COORDINATION

8.2.1 COORDINATION IN SAKHA

In Sakha, coordination of clauses can be expressed by proper coordinate constructions (syndetic as well as asyndetic), which are semantically as well as syntactically coordinate, as well as by pseudo-coordinate constructions (Haspelmath 2004), which are semantically coordinate but syntactically subordinate. Interclausal pseudo-coordinate constructions contain two predicates, one finite verb and one converb, making the converbal clause syntactically dependent on the finite clause. Pseudo-coordinate constructions formed with the sequential converb in *-An*, as exemplified in 8.9, are always same subject.

SAKHA

- (8.9) *Onu me:le iha:ri:la:n hie-bip-pit.*
 that.ACC simply fry.SQ.CV eat-PST.PTC-1PL
 'Those we simply fried and ate.'

(PIB: 173)

Despite the fact that this is probably the most common strategy to link clauses in both Sakha and Dolgan, it will not be discussed in detail here because the two languages apply this construction in an identical fashion.

In intraclausal coordination, pseudo-coordinate constructions in Sakha are formed with the postposition *kitta* ‘with’, but they are very rare. In fact, this construction is not attested in the corpus of narrative texts at all, but occurred only in elicitation.

SAKHA

- (8.10) *Uol-u kitta kī:s ki:ne-ye bar-al-lar*
 boy-ACC with girl cinema-DAT go-PRS.PTC-PL
 ‘The girl and the boy are going to the movies.’ (elicited)

In this construction the coordinate subject ‘the boy and girl’ consists of a syntactically asymmetrical construction of a noun *kī:s* ‘girl’, preceded by a modifying postpositional phrase *uolu kitta* ‘with the/a boy’. Formally, *kitta* behaves like a proper postposition in that it occurs after the noun *uolu* ‘boy’, and governs the accusative case. Thus, a word to word translation of 8.10 would be ‘the girl with the boy are going to the movies’. However, this is not the translation spontaneously given by Sakha speakers. Much more natural is the one given in 8.10 in which the boy and the girl are treated as semantically equivalent and thus as coordinate. This semantic equivalence is reflected morphosyntactically in the form of the verb *barallar*, which carries a third person plural suffix. This presupposes a third person plural subject, in agreement with the coordinate interpretation of the construction, and not a third person singular subject as would be the case for a comitative interpretation. While this pseudo-coordinate construction with *kitta* can be called at best marginal in Sakha spontaneous speech, it is included in the current discussion because it is gaining territory in Dolgan, as we will see below (8.2.3.3).

Proper coordination can be syndetic or asyndetic, and both strategies are used in interclausal as well as in intraclausal coordination². Syndetic coordination is expressed by coordinators such as *uonna* ‘and’ and *da* ‘and, but’, whereas

² While most morphological means can be used in interclausal, as well as in intraclausal coordination, intraclausal coordination has the additional option to link elements by means of suffixes of possession -*LA:χ* or comitative -*LI:n*.

asyndetic coordination is expressed through juxtaposition, whereby the nature of the connection is largely implicit in the semantics of the juxtaposed elements. Discourse context plays a major role in the disambiguation between different interpretations. Examples of syndetic and asyndetic coordination are given in (8.11) and (8.12).

SAKHA

- (8.11) *Kih̄in-īm min budduk mas erbi:-bin*
 winter-POSS.1SG 1SG this.like wood saw.SIM.CV-PRED.1SG
budduk, uonna mas budduk χajit-a-bin uonna
 this.like and wood this.like chop- SIM.CV-PRED.1SG and
kista:n ohox-χo ott-o-γun.
 bring.in.wood.SQ.CV stove-DAT heat- SIM.CV-PRED.2SG
 'In my winter I sawed wood like this, and chopped wood like this, and
 bringing in the wood you light the stove.' (ARR: 84)

- (8.12) *Min beh-is kila:s-ka İtik-Küöl-ge üören-ieχ-te:χ-pin,*
 1SG five-ORD class-DAT Y.K.-DAT learn-FUT.PTC-PROP-PRED.1SG
kiaχ suoχ, mama-m i:p-pat-a.
 opportunity NONEXIST mama-POSS.1SG send-NEG-PST.3SG
 'I should have gone to Ytyk-Kuol to the fifth grade, [but] there was no
 possibility, [therefore] my mother didn't send me.' (ARR: 58)

Regardless of the syntactic differences between these two strategies, from a semantic point of view syndetic and asyndetic coordination can express the same range of relations between conjuncts, including coordination, adversive coordination, causality and specification (Cheremisina 1995: 297). Only disjunction seems to be obligatorily expressed by an overt element. While asyndetic coordination is used more frequently, the remainder of this section will be concerned with syndetic coordination only. Dolgan and Sakha show no significant differences with respect to the use of asyndetic coordination, and therefore this clause linkage strategy is not relevant to the current discussion.

Morphologically the coordinators in Sakha can be divided into proper coordinators, fossilised verb forms, fossilised noun forms, fossilised adverbs and fixed expressions. Apart from these elements, which have a clear coordinative function, there is a group of modal elements that hold the middle between a modal and conjunctive meaning such as *töttörütün*, 'on the contrary', *χolobura* 'for

example' and several others. Table 8.2 gives a summary of the most common coordinators and coordinating expressions and the type of linkage they encode, based on information from grammars of Sakha (Cheremisina 1995, Ubryatova 1982: 472-474).

Table 8.2: Coordinative relations and their coordinators according to morphological type

	Coordination	Disjunction	Adversive coordination	Causality	Specification
Coordinator	<i>uonna, da, dayani, emie</i>	<i>bi:ter, du...du:</i>	<i>da</i>	<i>onon</i>	<i>bi:ter</i>
Fossilised (pro)noun	<i>itienne, onton, otton</i>		<i>otton</i>		
Fossilised verb	<i>buolan, buollayina</i>	<i>ebeter</i>	<i>buollayina, buolbakka, buolbatax, buollar, ebeter</i>		<i>ebeter</i>
Fixed expression			<i>ol ginan baran, ol ere:ri</i>	<i>ol ihin</i>	<i>ol ata</i>

8.2.2 COORDINATION IN DOLGAN

Coordination in Dolgan follows largely the same principles as Sakha. Asyndetic coordination is very common, and as far as syndetic coordination is concerned the majority of coordinators and coordinative expressions are also shared between the two languages. However, there are some striking differences, which primarily concern the frequency distribution of individual coordinators in Dolgan when compared to Sakha. An overview of the use of coordinators in the two languages is shown in figure 8.1. The bars represent the relative frequency of a particular coordinator with respect to the total number of coordinators in the corpus, which is 634 in Dolgan and 1323 in Sakha. Proportionally, this corresponds to 3.9% of all words in Dolgan and 4.5% in Sakha, which shows that the languages do not differ significantly with respect to the overall number of coordinators they use. The dark bars in Figure 8.1 represent the frequency for coordinators in Dolgan and the pale grey bars do the same for Sakha.

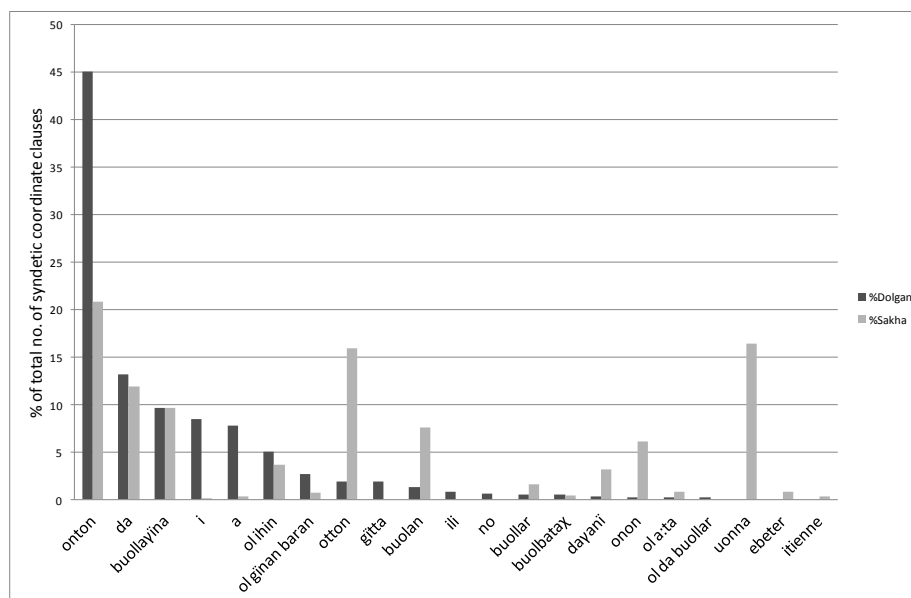


Figure 8.1 Relative frequencies of coordinators in Dolgan and Sakha

Most of the coordinators included in figure 8.1 overlap with the ones listed in Table 8.2 on the basis of the Sakha grammars. However, the match is not perfect. On the one hand, figure 8.1 includes the Russian coordinators *i* ‘and’, *a* ‘and, but’, *ili* ‘or’ and *no* ‘but’ in addition to the Sakha coordinators. These coordinators constitute 17.8% of all overtly marked coordinate structures in Dolgan speech, and in order to get an adequate impression of the overall number of coordinate clauses, they should not be omitted. Furthermore, the postposition *gitta* (Sakha *kitta*) ‘with’ is added to the list of coordinators. Although it is technically a postposition, it can be used in pseudo-coordinate constructions as well, as was discussed above (example 8.10). Section 8.2.3.3 will show that there is a difference in use and in frequency of this pseudo-coordinate construction in Dolgan and Sakha.

On the other hand, it appears that not all coordinators mentioned in Table 8.2 are relevant to the current discussion. For example, *bi:ter*, *buolbakka* and *ol ere:ri* do not occur in the corpora of Sakha and Dolgan at all and were therefore excluded from the analysis. *Emie* on the other hand occurs frequently in the corpus of Sakha and Dolgan, but only with the adverbial meaning of ‘also’ and ‘again’ and thus its coordinative use is not supported by the corpus data. Since inclusion of *emie* would misrepresent its function attested in spontaneous speech and since it would

erroneously raise the overall number of coordinate structures, this element was excluded from the analysis as well. Finally, some coordinative elements had to undergo a differentiated analysis before they could be included. This concerns multifunctional elements such as *da* or *onton*, which may have a coordinative function but may have other functions as well. For example, *da* can have the coordinative meaning of ‘and, but’, but it can also have the function of a negative particle or an indefinite particle. Similarly, the ablative demonstrative pronoun *on-ton* [ol-ABL, ‘from there’] has the function of a temporal coordinator with the meaning ‘and then’, but in other contexts it maintains its spatial meaning ‘from there’, in which case it must be analysed as an adverb. For such multifunctional elements their use was checked manually and only instances of clear coordinative use were included.

It needs to be noted that the preponderance of temporal coordinators such as *onton*, and the complete absence of other linking elements such as *bi:ter*, *buolbakka* and *ol ere:ri* is most plausibly explained by the discourse style that dominates the corpora. Most of the stories are oral histories, narrated as a monologue, in which temporal coordination (corresponding to English ‘and then’) naturally occupies a prominent place, and in which adversive coordination is naturally marginal because people do not normally contradict or disagree with themselves. Therefore, results could have been different had more dialogues been included, and more than one opinion been represented. However, in the currently available data these adversive coordinators only occurred in elicitation tasks.

Figure 8.1 shows a number of striking proportional differences between Dolgan and Sakha in the use of coordinative elements, four of which will be discussed in detail below: a) the complete absence of *uonna* in Dolgan; b) the highly frequent use of *onton* in Dolgan; c) the higher proportion of *gitta* with the semantic function of a coordinator d) the use of Russian coordinators in Dolgan discourse. Finally, there will be a brief note on the difference in range of actively used coordinative elements.

8.2.3 DIFFERENCES BETWEEN DOLGAN AND SAKHA

8.2.3.1 ABSENCE OF *UONNA* ‘AND’ IN DOLGAN

While most of the dissimilarities between Dolgan and Sakha are differences in relative frequency of coordinators and thus a matter of degree, there is one

absolute difference between the two languages, concerning the coordinator *uonna* ‘and’. While *uonna* is the second most frequent coordinator in Sakha and represents 16.4% of all overt coordinators in the corpus, this element does not occur in Dolgan at all. This is reflected by the lonely pale grey column for *uonna* in figure 8.1, which shows that it has no Dolgan counterpart at all.

According to Pekarskij, this coordinator has Turkic origins and is a contraction of the elements *ol *gĭnna* and is represented in Turkic as *sōnda* ‘after’ (Pekarskij [1907-1930] 1958-1959: 1839). Ubryatova, on the other hand, proposes that *uonna* developed from the sequence *ol kenne* [that after] (Ubryatova 1982: 98). While the exact origin remains unclear, both scholars agree that *uonna* can be traced back to Turkic origins. Thus it is plausible that its absence in Dolgan reflects the loss of this particular element in this language rather than an addition in Sakha after the languages began to diverge. This raises the question how a frequent element such as *uonna* could disappear from the language, and what alternatives today’s Dolgan employs to establish conjunctive coordinate relations. In order to explore these questions, it is necessary to get more detailed insights into the use of *uonna* in Sakha discourse. *Uonna* can be used in intraclausal (8.13) as well as interclausal coordination (8.14), in which case it can have a sequential aspect (English ‘and then’) as well.

SAKHA

- (8.13) *Bir saxa-li: kahiaččig-in, uonna mama-ta*
 one Yakut-ADVLR vest-ACC.3SG and mama-POSS.3SG
bier-bit solko kasinka-tin ial-lar-ga idždzi-bit-e [...]
 give-PST.PTC silk scarf-ACC.3SG family-PL-DAT carry-PST.PTC-POSS.3SG
 ‘One Yakut vest and the silk scarf her mother had given her she carried to
 the neighbours [...].’ (ARR: 033)

- (8.14) *E, papa-m otut toyus sil-la:χ-χa*
 hm papa-POSS.1SG thirty nine year-PROP-DAT
χa:ji-ttan taχχi-bit-a uonna onno
 lock.up.NLZR-ABL go.out-PST.PTC-POSS.3SG and there
öl-büt-e, ayjaχ ij buol-an bar-an.
 die-PST.PTC-POSS.3SG few month AUX-SQ.CV go-SQ.CV
 ‘Oh yes, and my father came out of jail in the year 39, and then he died after
 a few months.’ (ARR: 64)

Equivalent sentences from the Dolgan corpus show that these functions can be fulfilled by other native coordinative elements such as *da* 'and, but' (8.15) and *onton* 'and then' (8.16). Both *da* and *onton* are used mainly for interclausal coordination (8.15b, 8.16b), but are occasionally used for intraclausal coordination as well (8.15a and 8.16a).

DOLGAN

- (8.15a) *Vot tak pogib-lo voobs'e bu Khatangskij*
 PRT.R thus.R die-PST-N.SG.R in.general.R this Khatanga.ADJ.R
rajon taba-ta, Kheta da giene Katirik da giene
 district.R reindeer-POSS.3Sg Kheta PRT 3SG.POSS Katyryk PRT 3SG.POSS
 'Thus all the reindeer from the Khatanga district died, the ones in Kheta
 and the ones in Katyryk.' (APF: 85/86)

- (8.15b) *Ili: batta:-tilar da min interna:χ-χa χa:l-li-m buo*
 hand press-PST.3PL PRT 1SG boarding.school.R.DAT stay-PST-POSS.1SG PRT
 'They signed the forms [lit: pressed hands] and I stayed at the boarding
 school.' (NMC: 49)

- (8.16a) *Onton χatatsala:-bip-pit onton onno kim Polina Alekseevna ba:r*
 then ride.R-PST.PTC-1PL then there who Polina.R Alexeevna EXIST
e-te semja-tin gitta onton tjtja Nastja onton
 be -PST.3SG family.R-ACC.3SG with then auntie.R Nastja then
Ńuku:ska Ludmila Nikolaevna oyo-to onton
 Nicolay Ludmila Nicolaevna child-POSS.3SG then
Annuška tjtja Nastja
 Annushka auntie.R Nastja
 'And then we went on, there were P.A with her family, and auntie Nastja
 and Nukuska the child of L.N. and Annushka from auntie Nastya.' (DPK: 12)

- (8.16b) *it-tar-gin ba:j-a-gin onton köh-ö*
 dog-PL-ACC.2SG tie-SIM.CV-PRED.2SG then migrate-SIM.CV
tur-a-gin
 DUR-SIM.CV-PRED.2SG
 'You tie the dogs and then you migrate.' (IMA: 39/40)

It needs to be stressed that this use of *da* and *onton* is certainly not restricted to Dolgan. Table 8.2 above showed that these coordinators are used with the same functions in Sakha as well. However, figure 8.1 clearly illustrates that the frequency of occurrence of these two elements is higher in Dolgan. This applies in particular to *onton*, the proportion of which is more than twice as large as in Sakha. Against the background of the absence of *uonna*, this strikingly high proportion of *onton* in Dolgan could be interpreted as one way to cover the contexts in which Sakha would employ *uonna*. This line of thought is encouraged by a functional difference in the use of *onton* between Dolgan and Sakha, which will be discussed in more detail in the next section. Whether the loss of *uonna* came first, triggering an increase of the use of *onton* as a compensation strategy, or whether an increased use of *onton* in discourse caused *uonna* to disappear is impossible to tell without access to diachronic data for Dolgan. Unfortunately, these are rare and do not go further back in time than the 1930's, and at that moment in time *uonna* had already disappeared from the language (Ubryatova 1985, Ubryatova and Alekseev 2000). Nonetheless, the similarity in function between *uonna* and *onton* suggests that a connection between the absence of the one and the prominent representation of the other is plausible.

Another replacement for *uonna* as interclausal coordinator may be found in the adverb *onno*, which is not included in the list of coordinators because its analysis is speculative. This locative form of the demonstrative pronoun *ol* occurs equally frequently in Dolgan and Sakha, and it has a locative meaning in space (8.17) as well as in time (8.18). Often it is not possible to clearly distinguish the two, as in example 8.14, where *onno* could refer to the time when the speaker's father died, or to the place.

SAKHA

- (8.17) *Uonna ol kergen-ine:n onno olor-ol-lor.*
 and that spouse-COM.3SG there sit-PRS.PTC-PL
 'And he and his wife live there.' (ARR:113)

- (8.18) *Tihī:nča toyus sü:s tüört uon bi:r sıl seri, onno min*
 thousand nine hundred four ten one year war there 1SG
uon bi:r-de:χ-pin, anı.
 ten one-PROP-PRED.1SG now
 '(In) the year 1941 war, now there I was 11.' (ARR: 53)

While this locative adverbial meaning of *onno* is shared between Dolgan and Sakha, it seems that the locative-temporal meaning in Dolgan has expanded to cover a sequential-temporal aspect as well. In other words, the meaning of 'at that moment in time, back then' has expanded to contexts of 'thereupon, and then' ³, making it very similar to the interclausal linking function of *uonna* in Sakha.

DOLGAN

- (8.19) *Onno* *di:-bin:* "oyo-lor, *haŋar-iŋ* *kerget-ter-ger*,
 then say.SIM.CV-PRED.1SG child-PL say-PRS.IMP.2PL family-PL-DAT.2SG
[...] bies kopejka-ta bier-dinner, kino:-ga kim
[...] five copeck.R-POSS.3SG give-PRS.IMP.3PL film.R-DAT who
bar-iaχ-χitiŋ" di:-bin buo
 go-FUT.PTC-ACC.2PL say.SIM.CV-PRED.1SG PRT
 'And then I say: Children, tell your parents, your mothers that they give you
 five copecks, in order for you to go to the cinema, I say.' (LKS: 176)

The use of *onno* as a replacement for the interclausal coordinative function of *uonna* in Sakha may be represented by the following sequence of semantic change:

location in time → sequence in time → sequential coordinator

This may have taken place as a result of semantic and conceptual contiguity, potentially reinforced by the phonological similarity of *uonna* and *onno*.

8.2.3.2 FREQUENT USE OF *ONTON* 'AND THEN' IN DOLGAN

As mentioned above, *onton* is by far the most frequent coordinator in Dolgan. While this is the case for Sakha as well, the relative frequency in Dolgan is twice as high as in Sakha. In Sakha *onton* accounts for 20.8% of all overt coordinators, whereas in Dolgan it covers 45.1%. In Section 8.2.3.1 it was suggested that *onton* in Dolgan may have expanded its applicability to cover all the functions of *uonna* in Sakha (i.e. additional intraclausal coordination, in which there is no temporal aspect to the coordination), which could be one of the motivations for its high frequency in Dolgan. The next point of interest is whether there is an explanation

³ Cf. also Russian *na etom* [on that.PREP 'and then'].

for why this element and not another has adopted this function. On the one hand one may argue that the degree of functional overlap between *uonna* and *onton* in Sakha is so strong that they became interchangeable in certain dialects over time, and eventually the most frequent element (*onton*) took over. A scenario like this provides a purely language-internal explanation, in which dialectal variation eventually leads to an established change, motivated by changes in the text frequency of certain elements. However, it does not satisfactorily account for the complete absence of *uonna*, nor does it give any explanation as to why this happened only in Dolgan and not in the northern Sakha dialects that resemble Dolgan closely in other respects. An alternative perspective on this matter is provided by consideration of data from the neighbouring language Evenki, which show that the main coordinator in this language is an exact structural and functional equivalent of Dolgan *onton*. This suggests a potential role for Evenki in the development of this difference between Dolgan and Sakha. As in Dolgan and Sakha, coordination in Evenki is expressed mostly asyndetically, but syndetic coordination is possible too. In such constructions, the most commonly used overt coordinating element is *taduk* 'and' (Boldyrev 2007: 886, Nedjalkov 1997: 87), which is the ablative form of the demonstrative pronoun *tar* 'this'. As Boldyrev describes it, *taduk* [*ta-duk*, this-ABL] can be used "to connect equivalent constituents of a sentence, or [to connect, E.S.] entire sentences" (Boldyrev 2007: 886)⁴. Both uses are illustrated in examples (8.20a) and (8.20b).

It will be remembered that a very similar situation holds for Dolgan. *Onton* [*ol-(t)ta*n, that-ABL] was described as the ablative of the demonstrative pronoun *ol* 'that', and apart from its literal demonstrative meaning 'from there' it is commonly used to conjoin equivalent constituents or clauses, as was exemplified in (8.16a) and (8.16b), which are repeated here for convenience (see also Artemiev 2001: 140 for more examples)

EVENKI

(8.20a) <i>Hekupchu-l</i>	<i>tyrgani-l</i>	<i>ta-duk</i>	<i>inginipchu-l</i>	<i>dolboni-l</i>
hot-PL	day-PL	DEM-ABL ⁵	cold-PL	night-PL
'Hot days and cold nights.'				(Nedjalkov 1997: 90)

⁴ "...употребляется для связи однородных членов предложения или целых предложений." (Boldyrev 2007: 886).

⁵ original gloss: 'and'.

- (8.20b) *Bejetken* *togo* *daga-du-n* *teget-chere-n* *ta-duk*
 boy fire near-DAT-3SG.POSS sit-PRS-3SG DEM-ABL
nginakin *daga-du-n* *bi-si-n*
 dog near-DAT-3SG.POSS be-PRS-3SG
 'The boy is sitting near the fire and his dog is nearby.' (Nedjalkov 1997: 88)

- (8.21a) *Onton* *χatatsala:-bip-pit* *onton onno kim Polina Alekseevna ba:r*
 then ride.R-PST.PTC-1PL then there who Polina.R Alexeevna EXIST
e-te *semja-tin* *gitta* ***onton*** *tjotja Nastja onton*
 be -PST.3SG family.R-ACC.3SG with then auntie.R Nastja then
Ńuku:ska *Ludmila Nikolaevna* *oyo-to* ***onton***
 Nicolay Ludmila Nicolaevna child-POSS.3SG then
Annuška *tjotja* *Nastja*
 Annushka auntie.R Nastja
 'And then we went on, there were P.A with her family, and auntie Nastja and Nukuska the child of L.N. and Annushka from auntie Nastya.' (DPK: 12)

- (8.21b) *it-tar-gin* *baj-a-gin* ***onton*** *köh-ö*
 dog-PL-ACC.2SG tie-SIM.CV-PRED.2SG then migrate-SIM.CV
tur-a-gin
 DUR-SIM.CV-PRED.2SG
 'You tie the dogs and then you migrate.' (IMA: 39/40)

In addition, neither Evenki nor Dolgan have a specialised coordinative element that corresponds to *uonna* in Sakha. However, it needs to be mentioned that *uonna* does occur in other northern dialects of Sakha that did not have such close contact with Evenks as did Dolgan. Thus we observe a situation in which Dolgan and Evenki use a coordinative element that is different in form (*onton* vs. *taduk*), but is identical in morphological structure and in function (ablative demonstrative used in inter- and intraclausal coordination). This suggests that Dolgan may have assimilated to its unrelated neighbour Evenki, which resulted in the difference from its related neighbour Sakha that we currently observe.

If this is what happens, the most probable explanation for the structural transfer from Evenki to Dolgan is through the process of imposition by L1 Evenki speakers who learned Dolgan as a second language. As was explained in Section 3.1.3.3, structural transfer from L1 to L2 through interlingual identification is common during the process of imposition, whereby changes take place due to

extensive, but not perfect, structural and functional overlap of the comparable elements.

Recalling the historical and genetic data discussed in Chapter 2, the history of the Dolgan is characterised by a setting in which Sakha/Dolgan people and Evenks were in close contact, and where various degrees of bilingualism can be assumed with reasonable certainty in the Evenk and Sakha/Dolgan communities on the Taimyr Peninsula. However, due to the rise of Sakha as a lingua franca on the Taimyr during the 18th and 19th centuries, the Evenks who participated in the trade along the Khatanga Trading Way had to learn Sakha/Dolgan as an L2 rather than the other way round. In other words, that period of time may be distinguished by a considerable number of L1 Evenki speakers who learned Sakha/Dolgan as their L2. In addition, the large component of Tungusic genetic material in today's Dolgan population confirms historical sources mentioning close contact between the two ethnic groups. The large numbers of men and women of Evenki descent in the Dolgan population strongly suggest that certain groups of Evenks underwent a language shift, as a result of involvement in the activities along the Khatanga Trading Way or as a corollary of marriage with Dolgans. Given this social setting, one can imagine a situation in which Evenks who were learning Sakha/Dolgan sensed structural similarity in coordination structures between their L1 and their L2: as in their first language, coordination is either expressed *asyndetically*, or *syndetically* by a range of coordinating elements, the most common one being *onton*. Coincidentally, this element shares many functional characteristics with the most frequent element in their L1 (Evenki) *taduk*, which is used as a demonstrative, as an adverb and as an interclausal coordinator. Through interlingual identification, the similarity between the elements is enhanced. This structural and functional overlap may also explain why *onton* and not *uonna*, which also occurs very frequently in Sakha but has a different morphological structure, was associated with Evenki *taduk*. During the next stage, the similarity in function may have led to a complete identification of the two elements in the L2 learner's mind, while glossing over the subtle difference that *onton* in Sakha is only used for interclausal and not for intraclausal coordination, whereas *taduk* in Evenki can be used for both. The extended use of *onton* in Dolgan in intraclausal coordination may have rendered an element such as *uonna* functionally redundant, which may be why this element was not incorporated in the L2 Sakha/Dolgan lexicon of the L1 Evenki speakers. A feeling of functional redundancy may have been reinforced by the absence of specialised

coordinators in Evenki, due to which the L2 speaker may not have been on the lookout for such an element and therefore paid less attention to its occurrence in Sakha speech. This combination of factors (the complete identification of *onton* and *taduk* and the absence of a specialised coordinator in Evenki) could thus provide a probable explanation for the dominance of the coordinator *onton* in syndetic coordination constructions in Dolgan, as well as for the absence of *uonna*.

8.2.3.3 THE USE OF *ĠITTA* ‘WITH’ IN INCLUSORY COORDINATION CONSTRUCTIONS

In Dolgan and in Sakha, the main function of *ġitta* ‘with’ (or *kitta*, as it is spelled in Sakha), is a postposition. As can be seen from examples 8.22 and 8.23, this postposition carries a comitative meaning, which is characterised by Arkhipov (2009) by the following criteria: a) the predicate is not repeated more than once (resp. *bultaspitim*, *ata:rsan bardim*), b) the individual participants making up the participant set are expressed separately (implicit 1SG and *hilġihittari dʒonu* in 8.22 respectively, and *min* and *ġinileri* in 8.23), c) the expressions denoting these participants differ in structural rank (resp. *dʒonu* is dependent on 1SG, *ġinileri* is dependent on *min*), which is most obviously reflected by the fact that the verb only agrees with the grammatical subject (e.g. the person agreement in *bultaspitim* in example 8.22 is 1SG, despite the fact that semantically the hunters are plural).

SAKHA:

- (8.22) *Oh, dʒe, bu kim hilġi-hit-tar-ġi, kũdʒa:s*
 oh well this who horse-AG.NLZR-PL-ACC old
dʒon-u kitta bul-ta-s-pit-im.
 people-ACC with catch-VR-RECP-PSTPT-POSS.1SG
 ‘Oh well, I hunted together with who, with the horse herders, with old people.’ (AICH: 167)

DOLGAN:

- (8.23) *Min ġiniler-i ġitta ata:r-s-an bar-dĩ-m, [...]*
 1.SG 3.PL-ACC with accompany-RECP-SQ.CV go-PST-POSS.1SG, [...]
 ‘I began to accompany them.’ (LKS: 29)

These criteria, which are primarily morphosyntactic in nature, point to an asymmetrical structure and therefore constructions that conform to them cannot

be classified as coordinate. However, as was shown in Section 8.2.1, *gitta* is also used in pseudo-coordinate constructions, which are syntactically subordinate but semantically coordinate. An example for Sakha was provided in 8.10 and is repeated here as 8.24, supplemented by examples for Dolgan in 8.25 and 8.26.

SAKHA

- (8.24) *Uol-u kitta kīs ki:ne-ye bar-al-lar*
 boy-ACC with girl cinema-DAT go-PRS.PTC-PL
 'The/a boy and girl are going to the movies.' (XLE: 392)

DOLGAN

- (8.25) *Ol otto üle-bit bari-ta bihiene balig-i gitta kī:l [...]*,
 that PRT work-1PL all-POSS.3SG our fish-ACC with wild.reindeer [...]
 'That is all our work: fish and wild reindeer [...]' (ANS 115)

- (8.26) *Maša Afonij dzaɣtar-in gitta hugun xomu-n-al-lar*
 Masha Afoniy woman-ACC.3SG with berry collect-RFL-PRS.PTC-3PL
 'Masha and Afoniy's wife are collecting berries.' (elicited)

In these examples, criterion c) for comitative constructions is not fulfilled. While there is morphological asymmetry between the two noun phrases, i.e. one occurs in the accusative and the other in the nominative, they are semantically symmetrical. In 8.25, the fish are not accompanying the reindeer or the other way round, and in 8.24, 8.25 and 8.26 the order of the two noun phrases could be reversed without changing the truth value of the proposition. More significantly, the predicates in 8.24 and 8.26 carry plural marking like they do in coordinate clauses, suggesting that the boy and the girl (8.24) and Masha and Afonij's wife (8.26) are semantically equivalent and thus symmetrical. Thus, the semantic properties of *gitta* in these constructions are strongly reminiscent of the semantics of a conjunctive coordinator.

A particularly clear illustration is the pair of examples (8.27a) and (8.27b) where approximately the same phrase is repeated by the speaker, but with different coordination strategies: in example 8.27a the noun phrases *biester tüörter* are connected through asyndetic coordination, and in 8.27b the same noun phrases are coordinated by the *gitta*-construction.

DOLGAN

- (8.27a) A *min di:-bin:* "Lju:ba-γa *kör, molodies, bieχ*
 and 1.SG say.SIM.CV-PRED.1SG Ljuba-DAT look well.done.R always
 bies-ter tüör-ter"
 five-PL four-PL
 'And I say: Look at good Ljuba, always fives and fours⁶.' (LKS: 255)

- (8.27b) *Ulaχan kī:h-īm dnevnig-in kör-üöm, bies-ter-i*
 big girl-POSS.1SG diary.R-ACC.3SG look-FUT.1SG five-PL-ACC
 gitta *tüör-ter*
 with four-PL
 'I look at the diary of my eldest daughter, all fives and fours...' (LKS: 244)

While these constructions are possible in Sakha as well, they occur more frequently in Dolgan. In Sakha they did not occur at all in spontaneous speech and the only example came from elicited material. In Dolgan, on the other hand, this type of construction constitutes 1% of coordinators in spontaneous texts. In addition to a higher frequency of use, Dolgan has expanded the coordinative use of this postposition by developing a particular subtype of pseudo-coordinate construction, the inclusory construction, the model for which was most probably provided by Russian.

In Russian, the preposition *s* 'with' corresponds to Dolgan and Sakha *gitta* in many respects. Like *gitta*, it establishes a relationship of accompaniment between noun phrases, as in 8.28.

RUSSIAN

- (8.28) *On ezdi-l na poezd-e so svoej sobakoj*
 3SG travel-PST.M.SG on train-PREP with RFL.INST.F.SG dog.INST.F.SG
 'He traveled by train with his dog.'

Further similarity is found in the fact that it can be used in pseudo-coordinate constructions, or coordinate comitative constructions as they are called by Arkhipov (2009: 234) as in 8.29. In this example, the given coordinate translation is more appropriate than the literal translation "Sasha with Masha went [PL] to the cinema".

⁶ The numbers refer to grades one gets in school, whereby five is the best grade and one the worst.

RUSSIAN

- (8.29) *Saša s Mašej pošli v kino*
 Sasha with Masha.INST go.PST.PL in cinema
 'Sasha and Masha went to the cinema.'

Another typical feature of this type of construction in Russian is that it is inclusory when the syntactic subject has a singular referent (i.e. first, second or third person singular) (Arkhipov 2009: 235). Inclusory means that one of the two noun phrases (the syntactic subject) already includes the referent of the second noun phrase and therefore occurs in the plural, despite the fact that it has a singular referent (see example 8.30)⁷. Thus *mī* 'we' in 8.30 includes the speaker and his brother, even though in a typical coordinated noun phrase the first element should refer to only one of the coordinated elements and not both. Strikingly, an exact equivalent of this structure is found in Dolgan (8.31) and is importantly not encountered in Sakha.

RUSSIAN

- (8.30) *Mī s brat-om lovi-li riibu u prichal-a*
 1.PL with brother-INST catch-PST.PL fish.ACC at jetty-GEN
 'My brother and I (lit.: we with my brother) were fishing at the jetty.'
 (Arkhipov 2009: 235)

DOLGAN

- (8.31) *Bi:r-de bihigi Regina-nī gitta hildz-ar e-ti-bit*
 one-MULT 1.PL Regina-ACC with go-PRS.PTC be-PST-1PL
 'Once Regina and I went for a walk.' (DPK: 1)

Thus we have seen that *gitta* in Dolgan and Sakha shares many functional properties with *s* in Russian. Both elements are used as an adposition (a post- and preposition respectively) with a comitative function 'with' and as a coordinator in pseudo-coordinate constructions. While Russian may have played a role in the

⁷ Of course, it could be that the same rule applies when the syntactic subject has a plural referent, that is, when the meaning of example 8.30 would be 'my brother and us were fishing at the jetty'. However, since in these cases the syntactic subject appears in the plural anyway, inclusion of the referent of the second noun phrase would have no effect on the number marking of the syntactic subject. Therefore pseudo-coordinate constructions are only noticeable when the syntactic subject has a singular referent.

increased use of *gitta* as a coordinator, its influence is particularly salient in the inclusory construction, since these are only found in Dolgan and Russian, and not in Sakha. The exact match in morphosyntactic structure between the construction in Dolgan and Russian suggests that this particular construction has been introduced into Dolgan on the model of Russian. It needs to be mentioned that this construction is not abundantly present in the Dolgan corpus. More precisely, I have three examples of it in spontaneous texts, one of them given by an old woman from the up-river village of Volochanka, and two of them given by children (9 and 14 years old) from the down-river village of Syndassko. Since this change is occurring in a community that is undergoing language shift, and thus where linguistic dominance is changing too, the process of how this use of *gitta* entered the Dolgan language is not straightforward. The older woman, who grew up with Dolgan as her dominant L1, could have adopted this construction as a result of structural borrowing from Russian due to intense contact with this language. However, since language shift to Russian in her village of Volochanka is almost completed, it is more likely that her dominant, or most activated, language has now become Russian, which favours an explanation in terms of imposition instead (see Section 3.1.3.3). The young children who use the inclusory construction are also dominant in Dolgan in their pre-school years because they are growing up in the linguistically most conservative village of Syndassko. However, after a few years of schooling in Russian they are now perfectly bilingual and it is not clear which language is their dominant one. Based on my own observations in the village I am inclined to think that these children are balanced bilinguals, since they employ both languages with equal ease and proficiency. This may nonetheless lead to interference from the one language in the other, but the underlying process of borrowing or imposition is hard to define due to the absence of a clearly dominant language. Thus, while the process of change in these children cannot unambiguously be determined, for the older woman in Volochanka the process should rather be defined as imposition than as structural borrowing. To eliminate confusion, this is of course not imposition of Russian structures by Russians who shift to Dolgan, but by the Dolgans themselves who have become dominant in Russian and project structural properties of this language onto Dolgan.

This argumentation is reinforced by the fact that inclusory constructions have been adopted in other non-Slavic languages of the Russian Federation, where Russian has become dominant in the community. For instance, example 8.32 from

the Ersha dialect of Mordovian shows a comitative relative clause, in which the subject of the relative clause appears in the plural, even though it refers to the first person singular.

MORDOVIAN (ERSHA)

- (8.32) *Tan'e-s' [kona-n't' marhta min' vihse tonafn'-i-me]*
 tanja-DEF.NOM which-DEF.GEN with we together study-PST-1PL
tus' Mosko-w
 go.PST.3.SG Moscow-LAT
 'Tanja, with whom I (or we) went to school, went to Moscow.'⁸

(Aralova: fielddata 2007)

Thus, this section has shown that the use of *gitta* as a conjunction in pseudo-coordinate constructions has expanded in Dolgan when compared to Sakha. While this use of *gitta* is possible in Sakha, the more common usage of the postposition in this capacity in an area of intense contact with Russians and widespread bilingualism, suggests that this increase may have been motivated by contact with Russian. This contact-influence is particularly salient in appearance of the inclusory construction in Dolgan, in which *gitta* is also used as a conjunction, and which occurs in Dolgan only. Since this construction is characteristic of Russian, the use of *gitta* in this way is very likely to have developed through contact, more specifically as a result of imposition from the dominant language Russian onto Dolgan.

⁸ For this example, the Russian trigger sentence was:

- (8.33) *Tanja s kotoroj mĩ uchi-li-s' ueha-la v*
 Tanja with REL.PRON.INST.F.SG 1PL study-PST.PL-RFL leave-PST.F.SG in
Moskvu
 Moscow.ACC
 'Tanja with whom I (or: we) studied went to Moscow.'

This sentence in Russian is ambiguous with respect to the number of the syntactic subject. While evidence for contact-induced change would have been stronger in sentences in which the plural syntactic subject has an unambiguously singular referent, the fact that the Mordovian sentence is ambiguous in the same way as the sentence in Russian supports the hypothesis that they share the same underlying model.

8.2.3.4 RUSSIAN COORDINATORS IN DOLGAN DISCOURSE

The study of coordination strategies in Sakha and Dolgan also reveals a difference in the use of Russian coordinators. A comparison of the two corpora shows that Russian coordinators constitute a significant proportion of the conjunctions in Dolgan (17.8%), while they are virtually absent in Sakha. Of course, they do occur in the Sakha corpus as well in situations of code-switching, as is exemplified in 8.34, but these instances were not included in the analysis. After all, in such sentences the coordinators still figure in an entirely Russian context. They show no sign of incorporation into Sakha (or Dolgan), and therefore there is no reason to assume a change in this language. Therefore, for the current purpose only structures as in 8.35 were included, in which the Russian coordinator appears in an otherwise purely Dolgan (or Sakha) environment.

SAKHA

- (8.34) *Kiniler-i üčügej-dik uba:st-ï-bîn, üčügej-dik*
 3.PL-ACC good-ADVLZR respect-SIM.CV-PRED.1SG good-ADVLZR
kör-büt-üm, iḵ i poḵoroni-l-a,
 see-PST.PTC-POSS.1SG them and bury-PST-SG.F
spokojno oni sčas hiṭ-taḵ-tara di:.
 calmly they now lie-COND-POSS.3PL say.SIM.CV
 ‘[...] I respected them well, I looked after them well, and I buried them, so now they lie in peace.’ (ARR: 232)

DOLGAN

- (8.35) *Tugu da bil-bek-kit, i heme-li:-git*
 what.ACC PRT know-PRS.PTC.NEG-2PL and.R criticism-VBLZR.SIM.CV-2PL
 ‘You don’t know anything and you swear.’ (LKS: 283)

The two most popular Russian coordinators used in this capacity are *i* ‘and’ and *a* ‘and, but’, but *ili* ‘or’ and *no* ‘but’ are also used, albeit to a lesser extent. An overview of the comparative frequencies in Dolgan and Sakha is given in Table 8.3, in which the Russian coordinators are arranged by decreasing frequency in Dolgan. The table shows that the overall proportion of Russian coordinators in Dolgan is significantly higher than in Sakha. *i* and *a* occur with roughly the same frequency in the Dolgan corpus (8.5% and 7.9% of all coordinators respectively), whereas their presence in the Sakha corpus is negligible (0.08% and 0.3%

respectively). *ili* and *no* occur much less frequently in Dolgan (0.9% and 0.6% respectively, but still more than in Sakha, in which these coordinators do not occur at all.

Table 8.3: Frequency distribution of Russian coordinators in Sakha and Dolgan

	SAKHA		DOLGAN	
	No.	% of all coordinators.	No.	% of all coordinators.
<i>i</i> 'and'	1	0.08	54	8.5
<i>a</i> 'and, but'	4	0.3	50	7.9
<i>ili</i> 'or'	0	0	5	0.9
<i>no</i> 'but'	0	0	4	0.6
Total	5	0.4	113	17.8

Like the other changes in clause combining, these coordinators are used by all age groups, including the older people who are dominant in Dolgan as well as the younger people who are dominant in Russian. Since the shift is still ongoing and both groups are part of the Dolgan-speaking community, it is impossible to make a rigid distinction between the processes of borrowing and imposition to explain this instance of contact-induced change. Rather I would argue that both processes play a role in the development of these changes. The use of these overt coordinators is the result of borrowing in the people who are dominant speakers of Dolgan. Due to intense contact with Russian they borrow the substance as well as the structural consequences of these coordinators into their dominant Dolgan language. This mostly concerns the older generation, and certainly people older than 70. For the younger speakers, and in all probability people younger than 40, this change is the result of imposition, where their highly activated Russian language percolates through their use of Dolgan. Therefore, this is an instance of a linguistic change where the interplay of two different underlying processes results in the same linguistic outcome.

8.2.3.5 RANGE OF ACTIVELY USED COORDINATORS

Figure 8.1 showed the frequency distribution of coordinators in Dolgan and Sakha. The steep slope for the use of Dolgan coordinators is obvious from this figure, but the analytical eye may have spotted that the slope for Sakha looks rather different.

To make this clear for the average observer, the two slopes are represented in figures 8.2 and 8.3 below.

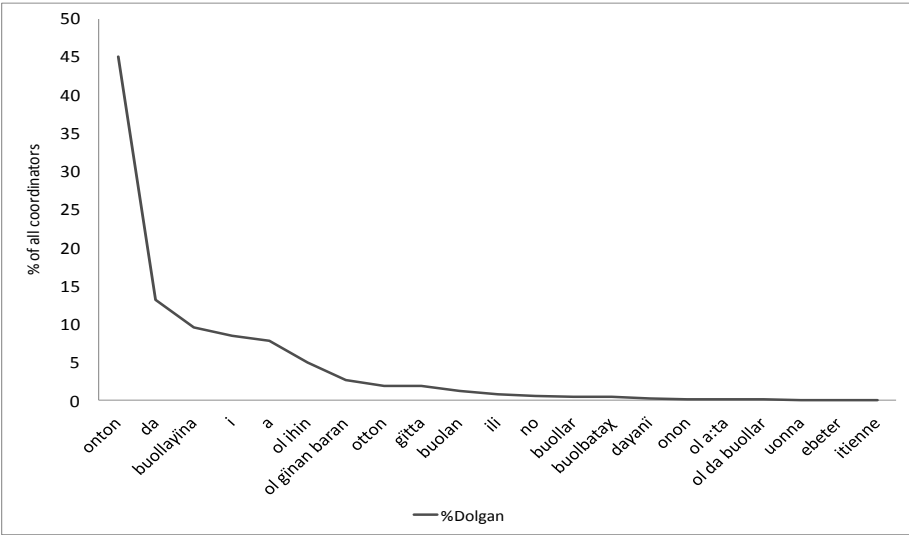


Figure 8.2 Coordinators in Dolgan in declining frequency of use

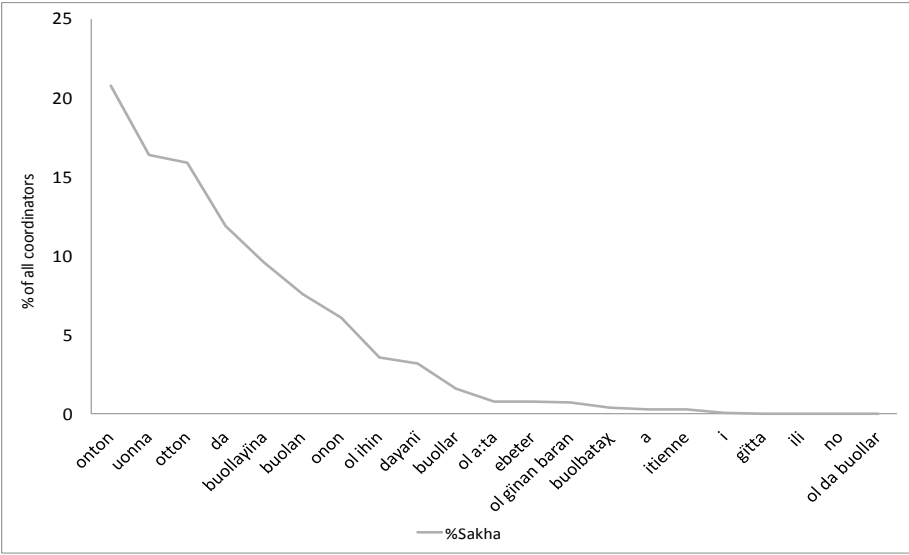


Figure 8.3 Coordinators in Sakha in declining frequency of use

The comparison shows that the slope in the diagram for Sakha goes down much more gradually than for Dolgan. The difference between the first and second most frequent coordinator in Dolgan is 31.9%, whereas in Sakha this is only 4.4%. Considering the fact that the overall frequency of coordinators in Dolgan and Sakha is comparable, this means that Dolgan has one coordinator, i.e. *onton*, which is used very actively, whereas the use of the others is comparatively limited. In Sakha on the other hand, the variety of actively used coordinators is larger. This shrinkage of diversity could be a sign of attrition as a result of the ongoing shift to Russian.

8.2.4 DISCUSSION

The previous sections have shown that syndetic coordination in Dolgan shows a number of differences from Sakha with respect to the presence, the frequency of use, and the function of certain coordinators. This contrasts with asyndetic coordination for which both languages behave identically. The main differences were identified as the absence of *uonna*, the high frequency of *onton*, the more frequent use of *gitta* in pseudo-coordinate constructions, in particular in inclusory constructions, and the presence of Russian coordinators in Dolgan discourse, whereby it was argued that the first two differences could be related. A comparison with Evenki showed that the most frequent conjunction in Dolgan (*onton*) has an identical morphological structure and functional distribution as the primary coordinate conjunction in Evenki (*taduk*), while its functional distribution deviates slightly from Sakha. It was suggested that this change in functional distribution (a spread from interclausal coordination to include intraclausal coordination as well) is potentially the result of a generalisation process in second language learning, which could be connected to the function of Sakha as a lingua franca, or eventually to shift. L1 Evenki speakers who learned Sakha/Dolgan may have projected coordination strategies from their L1 into their L2. The high degree of similarity in coordinate constructions between Evenki and Sakha/Dolgan may have facilitated this process and may have facilitated the loss of Sakha *uonna*, for which there is no equivalent in Evenki. Other elements in Dolgan that could fulfill the function of *uonna* are *onno*, possibly due to phonological similarity, and the Russian coordinator *i*, which is an exact functional equivalent. A more detailed discussion of the incorporation of Russian elements will follow in Section 8.4.

The increase in the use of *gitta* and the development of inclusory constructions was attributed to contact with Russian. It was argued that the inclusory construction, which is absent in Sakha, but standard in Russian, has been introduced into Dolgan through the process of imposition as a result of intense bilingualism and ongoing shift to Russian.

Thus, if the argumentation is correct, we can conclude that in coordination we find influence from both Evenki and from Russian, whereby Evenki has left its traces on Dolgan in the form of structural and functional change which is most probably the result of imposition and second language learning in a process of language shift from Evenki to Dolgan. Russian influence materialises in the form of changes in substance as well as in structure, which are introduced into the language as a result of borrowing as well as imposition. Always allowing for individual variation, the process of borrowing was typically associated with the generation over 70, which is dominant in Dolgan, and imposition was correlated with the age group younger than 40, which is most probably dominant in Russian. The age group in between is hard to classify due to large differences in linguistic dominance depending on the village in which the speakers grew up, the profession and attitude of their parents as well as their own aspirations.

8.3 SUBORDINATION

As was discussed in Section 8.1.2, subordination is characterised by asymmetry, which means that one of the clauses is cognitively or morphosyntactically dependent on the other. It was mentioned that a cross-linguistically valid categorisation of asymmetric relations appears to be a classification into adverbial relations, complement relations and relative relations (Cristofaro 2003: 39). In the context of differences between Dolgan and Sakha, only adverbial and relative relations will be discussed in detail, since both languages behave identically with respect to complement relations. For adverbial subordinate clauses it will be shown that differences between the languages are the result of direct influence from Russian, which is reflected by the introduction of Russian subordinators into Dolgan discourse. For relative clauses, the observed differences will be attributed to language attrition induced by ongoing language shift, and are thus an indirect consequence of contact with Russian.

8.3.1 ADVERBIAL SUBORDINATION

Adverbial relations are relations in which two States of Affairs (or propositions) are linked “such that one of them (the dependent SoA) corresponds to the circumstances under which the other one (the main SoA) takes place” (Cristofaro 2003: 155). Adverbial relations are further subdivided into relations of purpose, time, condition and reason (Cristofaro 2003, Givón 1990: 827-837, Kortmann 1997, Thompson and Longacre 1985). The current discussion deals only with purpose (8.3.1.1), temporal (8.3.1.2) and conditional relations (8.3.1.3). Reason will be discussed briefly in 8.3.1.4, but for this type of relation, differences between Dolgan and Sakha seem to be incidental and not the result of a systematic change.

8.3.1.1 PURPOSE

Purpose relations are defined as relations that “link two SoAs, one of which (the main one) is performed with the goal of obtaining the realization of the other one (the dependent one)” (Cristofaro 2003: 157). In Sakha this type of relation is expressed asyndetically as well as syndetically. Asyndetically purpose is expressed by case marked participles, the imperative or necessitative mood followed by the particle *dien*, or by converbs. Syndetically, purpose is expressed with the help of various postpositions. Participial purposive constructions typically contain the future participle on *-IAχ* with possessive case marking (8.36), but occasionally the present participle on *-Ar* is used as well. Both participles carry possessive case marking (dative or accusative), which may agree in person and number with the subject of the subordinate clause. According to Cheremisina (1995) participles are employed only in different subject subordinate clauses.

SAKHA

- (8.36) *Mannik* *hörü:n-ner-ge* *taba-ŋ* *üör-üŋ*
 in.this.way cool-PL-DAT reindeer-POSS.2SG herd -POSS.2SG
üčügej-dik *hinnán-an* [*ah-iay-in*] *örü:-gün.*
 good-ADVLZR relax-SQ.CV eat-FUT.PTC-ACC.3SG rest.one.day.SIM.CV-PRED.2SG
 ‘On cool days like this your reindeer relax well and you rest one day so that
 they can eat.’ (XKM: 17)

An additional frequent way of expressing purpose, which is not mentioned by Cheremisina but is in fact pan-Turkic, is the use of the multifunctional element *die-n* [say-SQ.CVB] in combination with the near future imperative (8.37). Alternatively it is combined with the necessitative mood, based on the future participle on *-IAχ* followed by the propriative suffix *-LAχ* and predicative person marking, as exemplified in 8.38. While clauses such as in 8.37 are used quite frequently and have clearly a purposive meaning, they sometimes occur on their own as well. This makes their subordinate status questionable, and suggests that in such contexts a desiderative interpretation may be more appropriate. However, in 8.37 there is clear cognitive dependency between the clauses and therefore this construction should be included in the category of proper purposive clauses.

SAKHA

- (8.37) *Oyonńor onnuk-ka üle-le:-bit, bili nú:čča-li:*
 old.man such.a-DAT work-VBLZR-PST.PTC that.one Russian-ADVLZR
haχa-li: bil-er buol-an nú:čča
 Sakha-ADVLZR know-PRS.PTC AUX-SQ.CV Russian
argıstas-tay-ına [kepse-t-tin di-en]
 accompany-COND-COND.3SG tell-CAUS-IMP.3SG say-SQ.CV
horuj-an ana:-bıt-tar.
 give.commission-SQ.CV appoint-PST.PTC-PL
 ‘The old man worked in such a one, since he knew Russian and Sakha they appointed him specially, so that he could talk (with them) when he accompanied Russians.’ (REX: 114)

- (8.38) *Ol ihin buolla:na ol mototsikl il-li-bıt,*
 that for however that motorcycle.R take-PST-1PL
onton hotoru [bult-uox-ta:χ di-en] anı ha:
 then soon catch-FUT.PTC-PROP say-SQ.CV now gun
il-li-bıt, ol kurduk dʒögüör-bütüger.
 take-PST-1PL that like Egor-DAT.1PL
 ‘So we bought the motorcycle, then soon after that we bought a gun so that he could hunt, so we did for our Egor.’ (XLE: 379)

Converbal purposive clauses can be formed with the sequential converb on *-An* (8.39) or with a special purposive converb on *-A:rI* (8.40). As the name says, the sequential converb in fact only encodes the sequence of two clauses but leaves the

nature of their relation unspecified. However, the relation can be interpreted as purposive if the semantics of the clauses allow for it (8.39). More specific and more frequent for this meaning is the use of the purposive converb on *-A:rl*. This converb may occur with predicative person marking agreeing with the subject, but this is not obligatory. Converbial purposive clauses are always same subject clauses.

SAKHA

- (8.39) *Onu kenniki manna [ostuoruja-tin il-an] bali:ha.*
 that.ACC afterwards here history.R-ACC.3SG take-SQ.CV hospital.R
arxi:ba-tin irit-tar-bip-pit tuoχ da huru-llu-bataχ.
 archive-ACC.3SG scrutinize-CAUS-PST.PTC-1PL what PRTwrite-PASS-PST.PTC.NEG
 ‘Afterwards in order to take his (medical) history we made the hospital
 archives scrutinize (everything), nothing was written.’ (REX: 126)

- (8.40) *Onon iye-m dze ol hordo-h-on bihigi*
 that.INST mother-POSS.1SG well that make.suffer-RFL-SQ.CV 1PL
i:t-en [abira-n-a:ri] ol čičimaχ-χa ülele:-bit-e.
 bring.up-SQ.CV help-RFL-PURP.CV that Chichimax-DAT work-PST.PTC-POSS.3SG
 ‘Therefore my mother suffered while bringing us up and worked in
 Chichimax in order to receive help.’ (PIB: 94)

According to the Sakha grammar, syndetic purposive clauses are formed with the help of the postpositions *tuhugar*, *ihin* and *innitten*, but Cheremisina admits that *ihin* and *innitten* occur very rarely, and that in the majority of cases these postpositions have the semantics of reason rather than purpose (Cheremisina 1995: 256). However, data from the spoken Sakha corpus do not give much support for the use of *tuhugar* in the function of purpose either. There are four instances of it in the corpus, but in all cases it serves to encode a beneficiary phrase, rather than a purposive clause, as is illustrated in 8.41 on the next page. Thus, data from the Sakha corpus show that the most common ways to encode purpose relations is by converbal constructions for same subject purposive clauses (converb on *-A:rl*), the imperative+*dien* construction for different subject purposive clauses where the subject is third person singular, or by participial constructions. The corpus does not provide evidence for the existence of syndetic purposive subordination constructions in Sakha, so if they do exist at all, they play at most a very marginal role.

SAKHA

- (8.41) *Onon* *če* *biligin* *ol* *oyo-lor-but* ***tuh-ugar***,
 that.INST well now that child-PL-1PL side-DAT.3SG
hien-ner-bit ***tuh-ugar*** *di-en* *bar-an* *hill-a-bit*.
 grandchild-PL-1PL side-DAT.3SG say-SQ.CV go-SQ.CV be-SIM.CV-1PL
 ‘So now we live for the benefit of our children, for the benefit of our
 grandchildren.’ (XLE: 520)

In Dolgan, purposive constructions are generally expressed with the same morphosyntactic means as described for Sakha. However, a number of differences must be noted. First, the range of possible constructions is slightly narrower, because Dolgan does not employ *dien* with a purposive meaning, making it an outlier in the Turkic language family (Matic and Pakendorf, in preparation). Second, more than half of the purposive clauses is formed with the help of the Russian purposive subordinator *štobi* ‘in order to’.

As in Sakha, the possessive-marked accusative form of the future participle expresses a purposive relation between main clause and subordinate clause (8.42), and the purposive converb on *-A:rl*, with or without predicative person marking, is productively used to this end as well, as exemplified in 8.43 and 8.44 respectively.

DOLGAN

- (8.42) [*Bi:r hir-ten* *nöjüö* *hir-ge* *dieri* *ti:j-ieg-in*] *onu*
 one place-ABL next place-DAT till reach-FUT.PTC-ACC.3SG that.ACC
di-e:čči-ler *turu:* *kuraŋ-a*, *ikki* *turu:*, *bi:r* *turu:*
 say-HAB-PRED.3PL post approximately-POSS.3SG two post one post
 ‘In order to reach the next place from the other they say approximately a
 turuu, one turuu, two turuu.’ (ANS: 53)

- (8.43) [*Dzie-ber* *köt-ö:rü*], *kürü:-bün* *buo*
 house-DAT.1SG fly-PURP.CV escape.SIM.CV-PRED.1SG PRT
 ‘In order to fly home I escape.’ (LKS: 38)

⁹ A *turu:* is a shamanic pole, and is used as a measure of distance.

- (8.44) *Kīaj-an* *u:ččax-ta-m-mat*
 be.able-SQ.CV riding.reindeer-VBLZR-RFL-PRS.PTC.NEG
er-dep-pititten *iti:* *iti:* *čexčeke* *kördö-n-ö*
 be-COND-ABL.1PL cry.SIM.CV cry.SIM.CV heap search-RFL-SIM.CV
hıldž-a:čči-bīt *u:čax-χa* *mi:n-e:ri-bit*
 be-HAB-1PL riding.reindeer-DAT mount-PURP.CV-1PL
 ‘From the moment we could barely ride a reindeer we looked for mounds
 while crying, in order to climb onto our reindeer.’ (PPK: 6)

In addition to these strategies, constructions with the Russian subordinator *štobī* ‘in order to’ occupy a prominent place in the formation of purposive relations. In Sakha, it does not occur at all, but in Dolgan more than half the purposive clauses (12 out of 23) are expressed with the help of this element.

In Russian, *štobī*-constructions consist of a main clause and a subordinate clause introduced by *štobī*. The predicate in the subordinate clause is either a past tense form (in different subject clauses, see 8.45) or an infinitive (in same subject clauses, see 8.45).

RUSSIAN

- (8.45) *On mne da-l den’gi, štobī ja kupi-la xleb*
 He 1SG.DAT give-PST.M.SG money, in.order.to 1SG buy-PST.F.SG bread.ACC
 ‘He gave me money, in order to buy bread.’

- (8.46) *On vzja-l den’gi, štobī kupit’ xleb*
 He take-PST.M.SG money, in.order.to buy.PST.INF bread.ACC
 ‘He gave me money, in order to buy bread.’

In Dolgan *štobī*-constructions, the form of the predicate typically is identical to the verb forms found in constructions without Russian influence: a future participle in possessive accusative case, agreeing with the subject of the subordinate clause. Thus, from a logical point of view one could argue that the addition of the Russian purposive conjunction is redundant and adds no semantic value to the construction. However, *štobī* does influence the overall structure of the sentence by increasing the flexibility of the position of the subordinate clause in the sentence. While the typical position of subordinate clauses with a typical Turkic structure is before the main clause, the typical position of sentences preceded by *štobī* is after the main clause, thus complying to the model provided by Russian

(although a sentence-initial position of such clauses in Russian is possible in certain contexts). Thus, while the element itself is semantically redundant, the addition of *štobī* has a structural effect on the organisation of main clauses and subordinate clauses in Dolgan.

DOLGAN

- (8.47) *I* *onu* *buollayina* *tur-uor-a-bīt* *buo*
 and.R that.ACC PRT stand-CAUS-SIM.CV-PST.PTC PRT
 štobī [*sīvorotka* *buol-uoy-un* *ke*]
 in.order.to whey.R become-FUT.PTC-ACC.3SG CONTR
 ‘And we put that away so that the serum appears.’ (IMA: 3)

8.3.1.2 TEMPORAL RELATIONS

Temporal relations involve the temporal sequence or simultaneity between a main proposition and a dependent one. They can be divided into relations of temporal posteriority, anteriority, and temporal overlap. In this classification, the terminology is based on the perspective of the proposition in the subordinate clause: temporal posteriority means that the proposition in the subordinate clause is posterior, or *follows*, the proposition in the main clause, and temporal anteriority that the subordinate proposition *precedes* the one in the main clause. Therefore, relations of temporal posteriority are, perhaps somewhat counterintuitively, also called ‘before’ relations, of temporal anteriority ‘after’ relations, and relations of temporal overlap are called ‘when’ relations (Givón 1990: 827-837, Cristofaro 2003: 159). As for purposive relations, Sakha and Dolgan have a diverse range of constructions to express temporal relations. These include, but are not limited to: a) sequential converbs on *-An* with optional predicative person marking to express anteriority (‘after’) as in 8.48) the future participle on *-IAχ* in the possessive marked dative case, followed by the postposition *dieri* ‘until’ to express posteriority (‘before’) as in 8.49, and in Dolgan simultaneity as well, as exemplified 8.50) the simultaneous converb on *-A* to express simultaneous events as in 8.51.

DOLGAN

- (8.48) *[ira:s-t-an bar-an-nar] kiptij-inan kirj-al-lar*
 clean-VBLZR-SQ.CVB go-SQ.CV-PRED.3PL scissors-INST cut-PRS.PTC-PRED.3PL
tü:-tün
 reindeer.fur-ACC.3SG
 ‘After cleaning, they cut the fur with scissors.’ (ESB: 6)

SAKHA

- (8.49) *Onton oskuola-ni büt-er-ieχ-per dieri*
 then school-ACC end-CAUS-FUT.PTC-DAT.1SG until
töhö bayar-ar interine:t, on-uh-u
 to.what.extent wish-PRS.PTC boarding.school ten-ORD-ACC
büt-er-ieχ-per da dieri min sa:s
 end-CAUS-FUT.PTC-DAT.1SG PRT until 1SG spring
sett-ih-inen toχto:-but-um, aχsī-h-inan.
 seven-ORD-INST stop-PST.PTC-POSS.1SG eight-ORD-INST
 ‘Then before I finished school boarding school as much as you like, even
 before I finished tenth (grade), I stopped in seventh, in eighth.’ (REX: 158)

DOLGAN

- (8.50) *Honon interna:χ-χa buol-lu-m,*
 that.way boarding.school.R-DAT be-PST-POSS.1SG
[ula:t-iaχ-par dieri] iti-ke:če:n hildzi-bīt-īm bu
 grow.up-FUT.PTC-DAT.1SG till this-ADVLR go-PST.PTC-POSS.1SG this
ñoχčo buol-an buo
 hunchbacked be-SQ.CV PRT
 ‘Thus I came to the boarding school, and while I grew up I became
 hunchbacked, and I lived like that.’ (NMC: 50)

DOLGAN

- (8.51) *[Hir-bitin kör-ö] hildz-a-bīt*
 earth-ACC.1PL look -SIM.CV AUX-SIM.CV-1PL
 ‘We travel while we look at our land.’ (ANS: 28)

As can be seen from these examples, the verb forms are non-finite, and the subordinate clauses are preposed to the main clause. The postposition *dieri* occupies the final position in the subordinate clause. A comparison of temporal

subordinate constructions in the Sakha and Dolgan corpus shows that there is only little difference between the two languages in this respect. Yet, there are two features which do not overlap. These are the use of the native postposition *dieri* ‘until’ and the use of the Russian subordinator *poka* ‘while’, which has found its way into the Dolgan language while it is not used in Sakha. Nonetheless, in contrast to purposive clauses, where more than half of the clauses are formed with a Russian element, Russian influence on temporal subordination, with six examples, is rather limited.

With respect to the first difference, in Sakha the only meaning of *dieri* is ‘until’, regardless of whether it is used in a spatial (8.52) or in a temporal sense (8.53).

SAKHA

- (8.52) *Bu u:-nu at-īnan tobug-ar dieri*
 this water-ACC horse-INSTR knee-DAT.POSS.3SG until
keh-erd-en ajan-nī:r-bīt
 wade-CAUS-SQ.CV journey-VBLZR.PRS.PTC-1PL
 ‘We travelled by making the horses wade up to [until E.S] their knees in this water.’
 (Uvarovskij: 243)

- (8.53) *Tudd-um mama-m kel-ier dieri*
 stand.PST-POSS.1SG mama-POSS.1SG come-FUT.PTC.DAT.3SG until
 ‘I stood until my mother came.’
 (ARR: 49)

In Dolgan this postposition has the additional meaning of ‘while’, or ‘as long as’ (or German ‘solange’ as in Stachowski 1993: 80) when it is used in the temporal sense. This use is not exceptional in Dolgan and an example was given in 8.50 above. This means that in Dolgan temporal sentences formed with this postposition can be either posterior (‘before’ or ‘until’) or simultaneous (while), whereas they are only posterior in Sakha. This distinction is important in the light of the second difference, the use of the Russian subordinating conjunction *poka* ‘while’ in Dolgan discourse.

In Russian, sentences formed with this subordinate conjunction are finite and the subordinate clause can occur before or after the main clause. The position of the obligatory *poka* is always sentence initial.

RUSSIAN

- (8.54) [Poka ja spa-la,] vor zalez v dom
 while 1.SG sleep-PST.F.SG thief climb in house
 'While I slept the thief climbed into the house.'

In Dolgan, *poka*-constructions are in principle identical to the temporal subordinate clauses that are not influenced by contact, except that they are preceded by the Russian conjunction *poka*. In other words, the predicate is non-finite, the clause-final postposition *dieri* is preserved, and even the head-final order seems to be preserved, unlike what is seen in the purposive clauses influenced by Russian. The only difference is the insertion of the Russian subordinator in clause-initial position.

DOLGAN

- (8.55) A iti uol [poka kör-üör dieri gini-ni]
 and this boy while.R see-FUT.PTC-DAT.3SG until 3.SG-ACC
 ta:s-χa iŋn-i-bit-te tūs-püt
 stone-DAT stumble-EP-PST.PTC-PST.3SG fall-PST.PTC
 'And that boy, while he was looking at her, stumbled over a stone.' (LSB: 15)

Thus, as with the purposive *štobi*-construction, the original construction is completely retained, and from a semantic point of view the addition of *poka* is redundant.

8.3.1.3 CONDITIONAL RELATIONS

In conditional subordination, differences between Dolgan and Sakha are very limited. The only point of divergence concerns the use of the Russian conditional subordinate conjunction. In Dolgan spontaneous speech, the Russian conditional subordinate conjunction *esli* 'if' is sometimes used in an otherwise Dolgan context, whereas it is not used in Sakha. However, such constructions are relatively rare, as only four out of 83 conditional clauses (4.8%) are formed with *esli*. As for the purposive and temporal adverbial clauses, the Russian subordinator does not seem to add or replace a semantic function, since it is followed by the conditional mood, which by itself expresses full conditionality. Its only function is again to make the position of the subordinate clause more flexible, in a similar fashion to what we

have seen for purposive clauses: instead of being restricted to the pre-main clause position, sentences beginning with *esli* can be put after the main clause as well, thus reflecting the flexibility of clause position in Russian.

DOLGAN

- (8.56) *Ma:ma, di:-bin heee, ʧajdaʧ dieb-ij ke [jesli*
 mother say.SIM.CV-PRED.1SG heee how say.FUT.1SG-Q PRT if.R
svatatsa-la:-taʧ-tarına, minigin kördö:-töʧ-törüne]
 ask.in.marriage.R-VBLZR-COND-COND.3PL 1.SG.ACC request-COND-COND.3PL
di:-bin buo
 say.SIM.CV-PRED.1SG PRT
 ‘Mother, what do I say if/when they come and ask me in marriage, and look
 for me?’ (APC: 9)

8.3.1.4 REASON

As was mentioned in 8.3.1, differences between Dolgan and Sakha in subordinate clauses of reason are unsystematic and incidental and they are negligible within the totality of other constructions that encode reason. However, for the sake of completeness, I will report two instances here, which are the only two examples in the corpus where the Russian subordinative constructions *potomu što* ‘because’ and *za to što* ‘for the fact that’, are used for this aim.

DOLGAN

- (8.57) *A Dolga:t-tar ehigi [potomu što Dolga:n-nar hir-der-iger*
 and Dolgan-PL 2.PL because.R Dolgan-PL earth-DAT.3PL
törö:-bük-küt], ol ihin Dolgan buol-uox-ʧut
 be.born-PST.PTC-2PL that for Dolgan become-FUT.PTC-2PL
 ‘And you are Dolgans because you were born on the earth of the Dolgans,
 therefore you become Dolgan.’ (LKS: 20)
- (8.58) [*Za to što kim, nu kömölös-püt ihin*]
 for that that.R who well help-PST.PTC for
 ‘Because he ehm, well, helped.’ (TIS: 12)

In a similar fashion as the constructions with *štobī* and *poka* and *jesli*, the Russian subordinators are redundant from a semantic point of view. Example 8.59 contains two clauses for reason, the first one being subordinate and introduced by the Russian *potomu što*, the second one being syntactically coordinate and introduced by *ol ihin*. The statement on redundancy is meant to apply to the first clause of reason only. Without *potomu što*, example 8.57 would be a paratactic structure with subordinate semantics for which the relation of reason must be inferred from the context, a construction commonly found in Dolgan. In 8.58 the postposition *ihin* expresses the same meaning as Russian *za to što* and is therefore redundant as well. However, since these constructions both occur in the corpus only once, I am inclined to treat them as nonce borrowings rather than as integrated constructions in the Dolgan language, with potential structural consequences.

8.3.2 RELATIVE RELATIONS

Relative clauses differ from adverbial subordinate clauses in that they exhibit no implicit semantic connection between the two linked propositions. Rather, “it is the speaker who arbitrarily selects two SoAs on the grounds that they share a participant” (Cristofaro 2003: 197). Traditionally, relative clauses are divided into restrictive and non-restrictive relative clauses. Both types involve two propositions, or State of Affairs in Cristofaro’s terms, of which the dependent one gives some specification about the main one (Cristofaro 2003: 195). The difference between them is that restrictive relative clauses restrict the set of referents, whereas non-restrictive relative clauses only provide additional information about the main clause referent without necessarily identifying it within a set of possible referents (Keenan 1985: 168-169). This is illustrated by the set of examples below, where (8.59) is a restrictive and (8.60) a non-restrictive relative clause (Cristofaro 2003: 195).

(8.59) We went to the Bach concerts [my friend got the tickets for]

(8.60) They went to a number of Bach concerts, for which they had booked tickets several months in advance.

According to Cristofaro, only restrictive relative clauses are subordinate. She supports her argument with evidence from sentence modification, such as negation, which in restrictive relative clauses affects both propositions, but in non-restrictive clauses only one, indicating that there is no semantic dependency relation between the two, which according to Cristofaro is a proof against their cognitive or semantic subordination.

As pointed out above, relative clauses are characterised by a shared participant, the so-called head of the relative construction, which plays a role in the main clause as well as in the subordinate clause (8.59 and 8.60). However, the range of roles this so-called head can play in the relative clause, and thus which syntactic and/or semantic functions are ‘accessible’ for relativisation, varies from language to language. This topic has been the focus of much research and has resulted in the so-called Accessibility Hierarchy (Keenan and Comrie 1977). This hierarchy, which is based on a thorough investigation of a sample of 49 languages, demonstrates that “languages exhibit certain constraints with respect to the syntactic roles that are accessible to relativization, and which strategies can be used for which roles” (Cristofaro 2003: 199). The original version looked as follows (Keenan and Comrie 1977: 66):

Subject > Direct Object > Indirect Object > Oblique > Genitive > Object of
Comparison

This hierarchy is implicational in the sense that if a particular syntactic function on this hierarchy can be relativised in a language, then all the functions to the left of it are accessible to relativisation, too. Despite the fact that the hierarchy has been debated and modified after its introduction, in particular with respect to the notion of subject and object, the general idea still seems to hold for many languages.

Another related way to classify languages typologically with respect to relativisation concerns the morphosyntactic encoding of the head noun in the relative clause. It appears that cross-linguistically there are four ways in which languages encode the head noun in the relative clause. Some languages use only one strategy, other languages may use more than one, in which case the question arises which functions of the Accessibility Hierarchy are encoded by which strategies. The strategies described by Comrie ([1981] 1989: 147) include: a) the non-reduction strategy, in which the head is represented identically in the main

clause and in the relative clause, appears in its usual position and carries its usual case marking; b) pronoun retention, in which the head noun is represented by a pronoun instead of a full noun in the relative clause; c) relative pronoun strategy, in which the head noun is expressed by a sentence initial pronoun, case marked for its syntactic function in the relative clause; d) the gap strategy, in which the head noun is not represented in the relative clause at all.

8.3.2.1 RELATIVE CLAUSES IN SAKHA

In Sakha, the typical relative clause is a preposed participial construction employing the gap strategy. There are no grammatical restrictions with respect to accessibility, which means that all functions can be relativised. This kind of construction conforms to the general profile of the Turkic language family, for which preposed non-finite relative clauses are the typical way to express relative relations (Pakendorf 2012), and also more widely to the profile of the proposed Siberian ‘linguistic area’ (Anderson 2006). This is not to say that there is no variation in relative clause formation within Siberia. In fact, there is considerable diversity, but this variation concerns agreement features between the non-finite verb form in the relative clause and the head noun, and not finiteness or the position of the clauses themselves (Pakendorf 2012: 257).

However, for Sakha a distinction must be made between subject and non-subject relative clauses. A subject relative clause is a construction in which the head noun has the function of subject in the relative clause, whereas in non-subject relative clauses the head noun occupies any function except subject. While in both clause types case marking of the head noun is determined by the predicate of the main clause (MC), they differ with respect to the representation of the relative clause-subject (RC-subject) in the MC.

In subject RC’s, the head noun has the same referent as the RC-subject and is not coreferenced in the MC in any way.

SAKHA

- (8.61) *Ol [tu:-le:γ-i bier-bit] oγonńor ep-pit*
 that fur-PROP-ACC give-PST.PTC old.man say-PST.PTC
 ‘The old man who had given the fur bearing (animals) said...’ (REX: 88)

In non-subject RC the referent of the head noun is different from the RC-subject, and in this construction the RC-subject is cross-referenced in the MC by means of possessive marking on the head noun. In example 8.62 this possessive marking can be seen in the gloss ACC.1SG, in which 1SG reflects the possessive marking.

SAKHA

- (8.62) *Onno tut-ar teril-ler-bin kör-dör-üöm onton*
 there hold-PRS.PTC equipment-PL-ACC.1SG look-CAUS-FUT.1SG then
 ‘I will show you my equipment that I use.’ (XKM: 11)

A possible motivation for this distinction is the fact that Sakha is a pro-drop language and often the subject is left unexpressed. This poses no problem for the identification of the RC subject in subject RC’s, because the RC-subject has the same referent as the overtly expressed head noun. However, ambiguity may arise when the RC-subject does not have the same referent as the head noun in the MC, as is the case in non-subject RC’s. In these cases, cross-referencing of person and number of the (pro-dropped) RC-subject on the head noun enables the hearer to identify the referent of the pro-dropped RC-subject more easily.

There is one exception to this rule. This is when the subject of the relative clause itself is marked with a possessive suffix that refers to the head noun (Pakendorf 2012: 272). For example, in kinship terms, the connection between two nouns is established by an *izafet* construction: the head noun is obligatorily marked with possessive marking agreeing in person and number with the modifier noun e.g. *učutal kergen-e* [teacher husband-POSS.3SG] ‘the teacher’s husband’. Now when the unmarked modifier noun (i.e. *učutal* ‘teacher’) is relativised, it does not receive possessive person marking.

SAKHA

- (8.63) *Bihigi kergen-e arax-s-an bar-büt učutal-ī*
 1PL husband_i-POSS.3SG_k divorce-RFL-SQ.CV go-PST.PTC teacher_k-ACC
 [*učutal-īn] tapti:-büt
 [*teacher-ACC.POSS.3SG_i] love.SIM.CV-1PL
 ‘We love the teacher whom her husband left.’ (Pakendorf 2012: 272)

Thus, it appears that if the possessive suffix on the subject of the relative clause is coreferential with the head noun, then possessive marking on the head noun is redundant and even ungrammatical, as is indicated by the form preceded by an

asterisk. From this we can conclude that the possessive connection controlled by the *izafet* construction overrules the possessive connection required by the relative clause.

8.3.2.2 RELATIVE CLAUSES IN DOLGAN

In principle, the rules for relative clause (RC) formation are the same in Dolgan and Sakha. For subject RC's both Dolgan and Sakha use a participial construction preceding the head noun, which carries no possessive marking if there are no explicit semantic reasons for it.

DOLGAN

- (8.64) *Ösüö iti ulaḡan-niḡ erej-de-n-er dzaḡtar*
 in.addition this big-ADVLZR torment-VBLZR-RFL-PRS.PTC woman
 ba:r buol-a:čč'i
 is.present be-HAB
 'Also there are women who give birth with many difficulties.' (APC: 216)

- (8.65) *Ol Nastja-ni iṭir-a:čč'i iṭ hīt-ar*
 that Nastja-ACC bite-HAB dog lie-PRS.PTC
 'There lies the dog that bit Nastya.' (Elicited)

However for non-subject RC's the match between Sakha and Dolgan is not perfect. In these constructions, Dolgan shows more variation and flexibility with respect to possessive marking on the head noun (obligatoriness and shape) (8.3.2.2.1); and morphosyntactic complexity (8.3.2.2.2). In this section I will illustrate these points with the help of examples from the corpus and elicitation tasks, and I will argue that the variation in Dolgan can be explained in terms of language attrition, cross-linguistic tendencies and differences in communication style between Dolgan and Sakha, which may be linked to the former function of Dolgan as a lingua franca.

8.3.2.2.1 POSSESSIVE MARKING

In most cases, possessive marking is used in the same way as in Sakha: in non-subject relative clauses, person and number of the RC-subject are cross-referenced

on the head noun as possessive marking. However, this rule is applied not as strictly as it is in Sakha. First, possessive marking does not always appear, and second, it does not always have the expected form.

The first point is illustrated in example 8.66. In this sentence, we have to do with a non-subject relative clause, in which *kihi* ‘person’ is the head noun, and the second person plural is the subject of the relative clause. According to the rules for relative clause formation in Sakha, we would expect second person plural possessive marking on *kihi*, yielding *kihi-gitin* [person-ACC.2PL]. In fact, we find this possessive marking in the Sakha translation of structurally comparable sentences, as can be seen in example 8.67. In Dolgan, however, we only find the non-possessive accusative case suffix *-ni* governed by the predicate *bayarabit* in the main clause and no possessive case marking on the head noun.

DOLGAN

- (8.66) *Bihigi ehigi Noskuo-tan kel-bit kihi-ni bayar-a-bit*
 1PL 2.PL Khatanga-ABL come-PST.PTC human-ACC love-SIM.CV-1PL
 ‘We love the man with whom you (PL) came from Khatanga.’ (Elicited)

SAKHA

- (8.67) *Ehigi Ba:taɣay-ttan birge massina-nan ayan-na:-n*
 2PL Batagaj-ABL one.COLL car-INST journey-VBLZR-SQ.CV
kel-bit uču:tal-gitin tapti:-bit.
 come-PST.PTC teacher-ACC.2PL love.SIM.CV-1PL
 ‘We love the teacher with whom you came from Batagaj by car.’ (Elicited)

However, this does not imply that possessive marking on the head noun would be ungrammatical. Speakers accept variants with possessive marking without hesitation, as can be seen from example (8.68). On my question whether the addition of possessive marking made any difference to the meaning of the sentence, the answer was that non-possessive *kihini* would mean ‘just a man’, whereas possessive marked *kihigitin* would mean ‘that specific man’, suggesting that possessive marking has to do with identification and specificity. However, it is unclear to what extent the speaker experiences a real difference between the two forms, and to what extent this distinction was invented on the spot to account in some way for the attested variation across speakers. In order to be sure, one would have to know what exactly was going on in the speaker’s head while she was uttering the sentence. Therefore it would be necessary to do targeted elicitation

on definiteness and specificity, but since Russian as the elicitation language, like Dolgan, does not have definite or indefinite articles to make this distinction explicit, even such elicitation tasks could not fully eliminate this uncertainty, and it would remain difficult to be certain how the speaker interpreted the input sentence.

DOLGAN

- (8.68) *Bihigi ehigi Noskuo-tan kel-bit kihi-gitin bayar-a-bīt*
 1PL 2.PL Khatanga-ABL come-PST.PTC human-ACC.2PL love-SIM.CV-1PL
 ‘We love the man with whom you came from Khatanga.’ (Elicited)

Relativisation of other functions shows even more variation. The possessor relative clause in 8.69a was initially given in this form, with non-possessive accusative marking on the head noun *dʒaxtar* ‘woman’. Upon inquiry whether possessive marked forms are possible as well, two more variants were given, one being *dʒaxtar-gin* [woman-ACC.2SG], which is the expected form according to Sakha grammar and where the head noun agrees with the RC-subject. However, the form *dʒaxtar-in* [woman-ACC.3SG], was given as well (8.69b), in which *-(t)in* is the suffix for the third person singular possessive, instead of the expected second person. This leads to the next point of discussion, namely of possessive marking that does not appear in the expected form, since the possessive marking in 8.69b does not agree with any constituent in the sentence.

DOLGAN

- (8.69a) *En untajka il-iax-xin bayar-a:čči dʒaxtar-ī*
 2.SG fur.boot take-FUT.PTC-ACC.2SG want-HAB woman-ACC
min lavka-ya kör-būt-üm
 1SG shop-DAT look-PST.PTC-POSS.1SG
 ‘In the shop I saw the woman whose fur boots you want to buy.’ (elicited)

- (8.69b) *En untajka il-iax-xin bayar-a:čči dʒaxtar-in*
 2.SG fur.boot take-FUT.PTC-ACC.2SG want-HAB woman-ACC.3SG
min lavka-ya kör-būt-üm
 1SG shop-DAT look-PST.PTC-POSS.1SG
 ‘In the shop I saw the woman whose fur boots you want to buy.’ (elicited)

DOLGAN

- An explanation for the emergence of such constructions could be the generalisation in function of the third person singular possessive marking, possibly motivated by the all-round presence of *izafet* constructions in Dolgan and Sakha, as described in Section 5.2.3.2. This construction encodes possessive relations between entities in the broadest sense of the word. In many cases a better description of its function would be the establishment of an association between objects, such as modifier-modified (8.71).

SAKHA

- In 8.71, the concept 'blacksmith' is expressed as a composite expression, in which the head noun *u:s* 'master' is modified by the modifying noun *timir* 'iron'. The connection between the two nouns is established by the possessive marking on the head noun *u:s*. Crucially, this possessive marking is the third person singular. This applies to all such constructions in which two common nouns are involved, thus yielding a high text frequency of third person possessive suffixes, the function of which is to simply to link two entities, rather than being associated literally with a

third person possessive meaning. Thus, it is possible that the marker *-(t)A* is acquiring an additional connotation of general association between elements, instead of only representing a third person singular.

There are a number of factors that may underlie such a development. First of all, as in many languages, relative clauses are rather uncommon in Dolgan spontaneous speech. It is much more common to express such propositions by multiple paratactic clauses, as will be elaborated in the next section. This holds in particular for functions low on the Accessibility Hierarchy, such as possessors. The cognitive complexity of such constructions and their related infrequency of use may be the reason why speakers are uncertain about the formation of such relative clauses. This is reflected in the attested variation and in the use of possessive marking that reflects general association rather than specific relations. Second, it may be a reflection of language attrition. Simplification and the loss or modification of infrequent structures through processes such as rule generalisation and meaning extension are often associated with attrition and this use of the third person singular would be an example. Third, the idea that we have to do with generalisation of the third person singular is supported by the cross-linguistic tendency to treat this person as the cognitively and grammatically unmarked category.

8.3.2.2.2 SYNTACTIC COMPLEXITY

The final point of differentiation between Dolgan and Sakha is the observation that Dolgan speakers prefer the use of paratactic structures to express complex propositions (including relative relations) over syntactically complex structures that are common in Sakha. Syntactically complex relative clauses constructions exist in both Dolgan and Sakha, but they are more frequent in Sakha than in Dolgan discourse. Although exact percentages are hard to give due to possible ambiguities of interpretation, a rough estimate shows that in the spoken corpus of Dolgan 0.8% of the total number of clauses is a syntactically complex relative clause (14 out of 1868 clauses), whereas in the Sakha corpus this is 4.7% (171 out of 3617 clauses).

As mentioned earlier, in both languages there are technically no restrictions with respect to the syntactic functions that can be relativised. Assuming that the Dolgan and Sakha people do not differ cognitively with respect to the number of

complex propositions they intend to express, the lower proportion of relative clauses in Dolgan indicates that Dolgan speakers choose different means to express these complex propositions. In addition, if relative clauses are used, there is a clear preference to use them for subject and object relativisation, i.e. the two functions highest on the Accessibility Hierarchy and not for lower ranked functions: in the Dolgan corpus, of 14 syntactically complex relative clauses, seven are subject relativisation, six object relativisation and one relativisation of location.

In elicitation tasks, complex propositions were sometimes expressed by syntactically complex constructions on request, but typically paratactic constructions were given as an initial response. Complex constructions were given only on further inquiry. Conversely, when presented with syntactically complex constructions, Dolgan speakers always accepted them without hesitation, which indicates that such constructions are grammatical in Dolgan, and certainly belong to the speakers' passive knowledge. However, in active speech production their use is very limited and they are outranked by paratactic constructions, which is shown in the elicited examples below. In these examples, first the target sentence is given for the relativisation of direct object (8.72), indirect object (8.73) and possessor (8.74). These targeted sentences are followed by the responses in Sakha and Dolgan, which clearly show the different preferences in the encoding of such complex propositions across the two languages: in all three cases Sakha uses preposed, embedded relative clauses, whereas in Dolgan the complex proposition is broken up into two paratactic clauses.

DIRECT OBJECT: *'On the chair the cat is sleeping, whom the Alexeevs chased out of the house.'*

SAKHA

(8.72a) *Kiriehile-ye [A.-tar ü:r-büt] kuoska-lara utuj-a sīt-ar*
 armchair.R-DAT A.-PL chase-PST.PTC cat-POSS.3PL sleep-SIM.CV lie.PRS.PTC
'On the chair the cat is sleeping, whom the Alexeevs chased out of the
house.' (elicited)

DOLGAN

(8.72b) *[Kreslo-ya utuj-a hīt-ar koška], [gini-ni Alekseev-tar*
 armchair.R-DAT sleep-SIM.CV lie-PRS.PTC cat 3.SG-ACC Alexeev-PL
dzie-tten bap-pīt-tar]
 house-ABL chase-PST.PTC-PRED.3PL
'The cat is sleeping on the chair, the Alexeevs chased him out of the house.'

(elicited)

INDIRECT OBJECT: *'This is the teacher to whom they gave a flat near the club.'*

SAKHA:

(8.73a) Bu [kulu:p tah-igar jie bier-bit] uču:tal-lara.
 DEM club.R outside-DAT.3SG house give-PST.PTC teacher.R-POSS.3PL
'This is the teacher to whom they gave a flat near the club.'

DOLGAN:

(8.73b) Bu dʒie-ni klub iksa-tigar iti učital-ga bier-bit-ter
 this house-ACC club close-DAT.3SG this teacher-DAT give-PST.PTC-PRED.3PL
'This house near the club, they gave it to the teacher.'

(elicited)

POSSESSOR: *'That is the woman whose house we will buy.'*

SAKHA:

(8.74a) Bu [bihigi jie-tin il-iaχ-ta:χ] jaχtar-bit.
 DEM 1PL house-ACC.3SG take-FUT.PTC-PROP woman-1PL
'That is the woman whose house we will buy.'

DOLGAN:

(8.74b) [Bu ba:r dʒaχtar] [bihigi gini-tten die-tin il-iaχ-pit]
 this EXIST woman 1PL 3.SG-ABL house-ACC.3SG take-FUT.PTC-1PL
'This woman here, we will buy a house from her.'

(elicited)

This syntactic simplification does not only apply to relative clauses in Dolgan, but may be a more general feature of communication style. An impressionistic comparison of narratives in Dolgan and Sakha suggests that in general sentences in Dolgan are shorter and morphosyntactically less complex than in Sakha. While space does not allow me to go into the discussion about what linguistic complexity is (but see Sampson, Gil and Trudgill (2009) for an elaborate treatment of the topic) on the surface it is clear that Dolgan discourse contains less converbial and relative clauses than Sakha. While this kind of simplification may be attributed to ongoing language shift and concurrent language attrition, this communication style may also have older origins. After all, we know that in the 18th and 19th centuries Dolgan was the lingua franca on the Taimyr (see Sections 2.4.2 and 2.5.1 and 2.5.2). It was

the language used by different ethnic groups for intergroup communication, which means that there were naturally many second language learners. As was discussed in Chapter 3, this exocentric language use often coincides with a higher degree of transparency and simpler morphological and syntactic structures than endocentric language use. Thus, the overall preference of paratactic structures over subordinate structures could be a consequence of ongoing attrition, but perhaps more plausibly, it could also be the corollary of a communication style that developed when Dolgan fulfilled the role of intergroup language.

8.4 THE USE OF RUSSIAN LINKING ELEMENTS

The analysis above has highlighted the main differences in clause combining between Dolgan and Sakha, and it has shown that contact with Evenki, as well as with Russian has played an important role in the development of these differences. In this section a more in-depth discussion of the use of Russian coordinators will be provided and will be put in a cross-linguistic perspective.

Throughout the chapter it will have been observed that Russian influence is pervasive in both coordination and subordination, and a numerical confirmation of this impression is given in Tables 8.4 and 8.5. Since coordination and subordination are expressed by different means in Dolgan, two different tables are given to evaluate the significance of the Russian linking elements in a sensible way. For coordination the proportion of Russian coordinators is shown relative to the total number of coordinators within the conjunctive, adversative and disjunctive categories. For subordination such comparison was impossible due to the absence of overt native elements to encode subordination. Therefore for this category the proportion of Russian coordinators is calculated relative to the total number of purposive and conditional sentences in the corpus. Temporal relations were excluded from this comparison, because too often there is no overt marking at all that provides evidence of a temporal relation, as in the widespread use of converbial clauses. Therefore, the 6 instances of *poka* cannot be evaluated in percentages, but considering the high number of converbial clauses in the corpus, its share in the encoding of temporal relations can be confidently said to be very low.

Table 8.4: Proportion of Russian coordinators per overtly expressed coordination category

Coordinator	Russian	No.	Total no. of coord. in category	% Russian coord. per category
Coordinative	<i>i</i>	54	393	13.7%
Adversative	<i>a, no</i>	50, 4	205	26.3%
Disjunctive	<i>ili</i>	5	5	100%

Table 8.5: Proportion of Russian subordinators per subordinated clause type

Subordinator	Russian	No.	Total no. of sent. in category	% of Russian subord. per category
Purpose	<i>štobi</i>	12	23	52%
Conditional	<i>esli</i>	4	83	4.8%

The overt Russian elements are very obviously present in coordination, but also in the domain of subordination, except for relative clauses. This situation is not unique at all from a cross-linguistic point of view. In fact a significant amount of literature is devoted to the question why it could be that this type of linguistic material, in particular coordinators, is so accessible or ‘vulnerable’ (Matras 1998: 281) to copying. In this section a brief overview is given of some significant ideas on this matter, and it will be evaluated which approach could be most relevant for an interpretation of the Dolgan data.

Early accounts dealing with the transfer of conjunctions from one language to another emphasise above all the importance of structural properties of the languages in contact and of the copied linguistic element. One claim is that foreign grammatical elements, including conjunctions, are copied to fill a ‘grammatical gap’ in the recipient language (e.g. Heath 1978: 115-116, Campbell 1987: 279 and implicit in Mithun 1980: 96). In other words, the copied conjunctions are an overt expression of grammatical relations that were not explicitly encoded before. Supposedly, they are perceived as a useful addition to the existing grammatical system by speakers of the recipient language, which would be the reason why they are easily adopted. In addition, the morphosyntactic structure of the element itself is thought to influence the ease with which it is copied. The hypothesis is that the more an element is integrated into the morphological structure of the language, the less likely it is to be copied (Weinreich 1953: 41, Heath 1978: 72, Aitchinson 1981: 120), i.e. a suffix would be less likely to find its way into another language

than an unbound morpheme. A similar approach is taken by Moravcsik (1978). She proposes a number of implicational hierarchies which are based on the assumption that items with structural autonomy and referential stability are more likely to be copied in an early stage of contact than items without these characteristics (Moravcsik 1978: 110-113) and since conjunctions apparently are often unbound elements they fall into that category¹⁰. Thomason and Kaufman do not emphasise the importance of structural properties as much as the previous two accounts, but they do mention conjunctions as one of the most easily transferred grammatical elements in contact situations. They assign this kind of transfer to the category of “slightly more intense contact” (Thomason and Kaufman 1988: 74), which is the second out of five levels of contact intensity, and the first in which grammatical transfer occurs at all. Thus, this line of thought (with the exception of Thomason and Kaufman) links the ease of copying directly to the structural properties of the languages in contact: copied conjunctions either fill a structural gap or they are unbound elements.

However, there is clear counterevidence to such a purely structural account. For example, Stolz and Stolz (1996: 102-1023) provide data for languages from Mesoamerica, which show that copied clause linking elements do not always fill a structural gap. They show that Spanish conjunctions were copied into Mesoamerican languages, despite the fact that they already had explicit ways to express clause linkage. Moreover, the context in which the Spanish elements occur shows that they have not supplanted the indigenous morpheme, but that they can co-occur with it, even within the same construction, as illustrated in 8.75. This goes against another postulation expressed in the literature that a copied element always replaces an indigenous strategy (Weinreich 1953: 31-37, Heath 1978: 72).

ZOQUE

(8.75) *Si* *'izin* *is-pa-pi't* *te'y* *machete 'in* *ce'koŋ-pa*
 when PRO.1.SG:EMP see-INC-when PSR.3.SG machete PRO.1SG ask-INC
 ‘When I see him, I will ask him for his machete.’

(Knudson 1980: 139, in Stolz and Stolz 1996: 103, translation mine)

¹⁰ But note Matras' comment that this conclusion may be due to the biased dataset that is used in studies of conjunction copying. Most studies on this topic are based on contact between a recipient language and Spanish or Arabic as a model language, which both happen to have morphologically unbound conjunctive elements.

Stolz and Stolz relate the use of Spanish elements to the prestige Spanish enjoys in these communities and they suggest that the use of coordinative elements is an easy way for speakers to identify with this cultural group, while preserving most of their native language. Counterexamples to the criterion of structural integration have also been provided long ago. Heath himself (1978: 98-100) describes the case of Ritharngu, where a bound negative suffix from Ngandi has replaced the free negative morpheme that was native to the language.

Matras opposes to the structural approach and argues that the appearance of conjunctions at the top of the borrowability hierarchy is conditioned by cognitive properties, which “must be formulated in functional-communicative terms” instead (Matras 1998: 285). Rather than discussing conjunctions only, he talks about a category of utterance modifiers, in which he includes discourse regulating elements (including conjunctions), discourse markers and focus particles. In his view, the bilingual speaker has a higher mental processing demand than the monolingual speaker¹¹, which he tries to level out through convergence of the two language systems. Utterance modifiers

regulate linguistic-mental processing activities that can be attributed to what I call the “grammar of directing”. Bilinguals [...] are tempted to reduce the overt representation of the “grammar of directing” to just one set of elements. Preference is then given to the pragmatically dominant language. (Matras 1998: 291)

Thus, the use of utterance modifiers of just one set reduces the mental overload of the bilingual speaker. Which language system surfaces in the encoding of this grammar of directing is normally determined by the pragmatic dominance of the languages. Matras argues that the need for a functional instead of a purely structural motivation is highlighted by the fact that utterance modifiers with equal structural and syntactic status show different behaviour with respect to copying cross-linguistically. On the basis of data from Romani dialects, supported by a range of languages under Islamic and Spanish influence, he postulates the following borrowing hierarchy of coordinating conjunctions: but > or > and (Matras 1998: 303). In other words, elements equivalent to ‘but’ are copied before ‘or’, which are copied before equivalents to ‘and’. A discussion of Matras’ cognitive explanation for this hierarchy goes beyond the scope of this chapter, but the

¹¹ But see Section 3.1.3 for a critical view on this matter.

emergence of the hierarchy from his selection of languages is an interesting fact. While the available synchronic data from Dolgan cannot tell us anything with certainty about the order in which the coordinative elements were introduced into the language, the large differences in frequency are suggestive. Reasoning that more frequently used elements have become more established in a language than less frequently used ones, the data for Dolgan lead to the hypothesis that *i* 'and' was copied before *a* 'and, but', before *ili* 'or', before *no* 'but'. This would contradict the hierarchy proposed by Matras, but diachronic discourse data are needed to provide stronger support for this hypothesis.

With this theoretical background in mind, we can consider which factors are most relevant for the interpretation of the Dolgan data. Considering the omnipresence of asyndetic structures in coordination as well as in subordination in Dolgan and Sakha, an account relating the insertion of Russian elements to the filling of structural gaps may seem appealing at first. However, despite the dominance of asyndetic structures, the previous sections have shown that Sakha and Dolgan have a range of native coordinative elements, which partly overlap in function with the Russian ones. Therefore, the data for Dolgan do not support the idea that coordinators are copied to fill a structural gap: while coordinating elements are not obligatory, they do exist and are used frequently. Dolgan also does not lend support to the idea that copied linking elements always replace native elements, because the Russian elements occur in alternation with native elements, or sometimes even within the same construction (cf. Campbell 1993: 98, Stolz and Stolz 1996: 102-103). Rather than thinking of Russian elements as a replacement for the native ones, they could be seen as additions, the function of which overlaps more with the native element in some cases than in others. Thus, although the insertion of coordinative elements may be facilitated by the structural properties of both languages (the structural independence of the Russian morphemes and the optionality of such elements in Dolgan), it is unlikely that the main motivation for their appearance in Dolgan is to fill a structural gap. Stolz and Stolz (1996: 110) argue that communication style and prestige may be part of the explanation. As many dominant outgroup languages, Russian is associated with prestige, education and progress, what many people like to identify with and aim for, in particular for their children and which could be a reason why the use of Russian lexical items is favoured. However, while language structure and communication style may play a supporting role, the main reason behind the prominence of Russian coordinators, and to a lesser extent

subordinators, seems to be the dominance of Russian and the high degree of bilingualism in the Dolgan community due to the current shift to Russian. The undeniable relevance of these factors in a population that is in an ongoing process of shift suggests that a more appropriate explanation must be formulated in terms of psycho- and sociolinguistic processes that play a role in discourse organisation, which approaches most closely the functional-cognitive explanation proposed by Matras.

The examples have shown that Russian coordinators often occur in an otherwise Dolgan context. It could be the case that the linkage of events, and thus the structuring of discourse, indeed takes place on a different cognitive level than the organisation of a clause. This is what Matras calls the 'grammar of directing'. In some speakers, in particular the older ones, the use of Russian elements such as *i* 'and' and *a* 'and, but' may be an instance of borrowing due to the high text frequency of those elements in Russian discourse. In others, in particular speakers younger than 40, the use of Russian elements is the result of imposition. Although these people are Dolgan, their dominant language is Russian, and the constantly high activation of this language has caused them to adopt the linguistic and cognitive framework for relating events and organizing discourse that comes with the use of Russian language. Such a change in the organisation of discourse is in a way no more than an extension of structural imposition, where 'structure' now applies to the composition of discourse instead of morphosyntax alone. Thus it can be concluded that the case of Dolgan yields most support for a functional-cognitive explanation for the adoption of clause linking elements from Russian, whereas structural factors and issues of style and prestige may play a facilitating but subordinate role.

8.5 CONCLUSION

This chapter has shown that clause combining strategies in Dolgan and Sakha differ in a number of ways. While no clear differences occur in asyndetic clause linkage, syndetic strategies of clause combining show variation with respect to several features. The extended use of *onton* 'then' in Dolgan was argued to be attributable to contact with Evenki, the remaining changes were proposed to be caused by contact with Russian and the ongoing shift in Dolgan communities from Dolgan to Russian, or with the use of Sakha as a *lingua franca*. First this shift has

led to the transfer of overt coordinators and subordinators from Russian and to the introduction of structural models from Russian, such as the use of *gitta* and postposed subordinate clauses in Dolgan. Since these are all ongoing changes in a speech community that is changing quickly and in which there are significant differences across individuals with respect to linguistic dominance, it is not possible to explain the changes by referring to a single underlying process. Therefore it was argued that these changes are the result of both borrowing and imposition, dependent on the linguistic dominance of the speaker. Borrowing applies to those who are dominant in Dolgan and introduce these features due to intense contact with Russian, imposition to those Dolgans who have better command of Russian and project the structures of this dominant language onto Dolgan. Second, shift to Russian is the cause of ongoing language attrition, which surfaces in features such as decreasing diversity of actively used coordinators and simplification of morphosyntactic structures, which is particularly noticeable in the formation of relative clauses.

In addition to this shift-based account, the morphosyntactic simplification, as well as the general tendency in Dolgan to make shorter and more paratactic sentences when compared to Sakha, was explained by a difference in communication style. This way of speaking could have developed as a result of the function of Dolgan as an exoteric intergroup language to facilitate communication between different ethnic groups on the Taimyr Peninsula. The differences and their explanations are summarised in Table 8.6 below.

Table 8.6: Contact influence in Dolgan clause combining strategies

	EVENKI IMPOSITION Coord.	RUSSIAN			
		BORROWING/IMPOSITION		ATTRITION/EXOTERIC USE	
		Coord.	Subord.	Coord.	Subord.
SUBSTANCE		<i>i</i> <i>a</i>	<i>štobī</i> <i>poka</i> <i>(esli)</i> <i>(potomu što)</i> <i>(za to što)</i>	less diversity	less strict use of poss. marking
STRUCTURE	use of <i>onton</i>	<i>gitta</i> constr.	postposed subord. clauses		more parataxis

It is important to stress the fact that this classification of differences is not as strict as the format of a table might suggest. As was argued before, the same linguistic outcome may, and often does have multiple explanations, which cannot be teased apart. The balance of linguistic dominance is currently so much in motion in the Dolgan community that it would be artificial to try to do so. This applies to the Russian coordinating elements and their structural consequences, but also to a lesser extent to the construction with *gitta* and to the use of more paratactic constructions, which was proposed to be either the result of current language attrition or a reflection of the more ancient function of Dolgan as a lingua franca. Finally, the loss of *uonna* in Dolgan was associated with shifting speakers from Evenki, who projected their native coordination structures onto their newly acquired language during the process of second language learning.

To conclude, this survey of differences in clause combining between Dolgan and Sakha confirms both the historical contact with Evenki as well as the pervasiveness of the Russian language in Dolgan communities today, which is reflected by its influence on substance as well as on the structure of Dolgan. It needs to be mentioned here that further influence from the neighbouring indigenous languages Evenki and Nganasan can be excluded in this respect. First, many features of coordination and subordination constructions are similar in Dolgan and Sakha (Turkic), Evenki (Tungusic) and Nganasan (Samoyedic). For example, all these languages make extensive use of asyndetic coordinate constructions and of preposed converbs and participles for the formation of adverbial and relative subordinate constructions respectively (Nedjalkov 1997, Teresh'enko 1979). However, the main difference between these languages on the one hand and Dolgan on the other is that the subject of the relative clause is cross-referenced on the participle in the relative clause instead of on the head noun, and this feature is not found in Dolgan at all. Conversely, none of the attested changes in Dolgan (except the use of the ablative demonstrative as a coordinator) is a prominent feature of either Evenki or Nganasan, whereas many of them are in Russian.

9.1 DRAWING THE STRANDS TOGETHER

The previous chapters have presented a number of differences between Dolgan and Sakha that most probably arose through contact with Evenki on the one hand and with Russian on the other. Based on a comparison of Dolgan with other Turkic and Tungusic languages, as well as on the fact that Dolgan history is characterised by frequent contact with other ethnic groups, it was argued that these differences represent changes in Dolgan rather than in Sakha. The survey of the contact-induced changes, as well as the underlying processes shows a heterogeneous picture. Changes were described for the lexicon, morphology and morphosyntax of Dolgan, and some were associated with the process of borrowing, others with imposition and again for others both processes seemed to have played a role.

So far, the primary focus has been on the identification of individual contact-induced changes and on their analysis in terms of social and historical factors. However, in the quest for insights into which role contact-induced linguistic change can play in the reconstruction of a people's prehistory, it is necessary to extend our scope beyond a simple inventory of changes. For this purpose, we need

to zoom out from the individual differences and view the contact situation as a whole, including not only the linguistic information, but also material from socio-historical, ethnographic and genetic sources. Only by embedding the linguistic changes in this context, can we properly evaluate their significance.

Both socio-historical information and theories of language contact provide indispensable clues for understanding contact-induced changes as well as for reconstructing the social setting in which they may have occurred. However, the conclusions drawn on the basis of this information alone remain tentative. To be sure, the correlations between contact-induced changes and social settings that have been proposed in different theories of language contact are based on cross-linguistic generalisations over a wide variety of case studies and therefore carry a high probabilistic value. However, contact linguists themselves recognise that these generalisations are anything but absolute. There are simply too many variables in a contact situation to predict the linguistic outcome. The same holds for socio-historical information. As was pointed out in Section 2.4.2, it is impossible to acquire a complete picture of all the social factors that have played a role in the linguistic outcomes of a contact situation, because historical reports may be skewed by the writer's intentions or obligations, by limited access to the communities in question, or by chance. Therefore, on the basis of these two kinds of data, we can only deduce likely scenarios in which the changes occurred.

The only kind of information that is not influenced by peoples' opinions and intentions is our genetic material. As explained in Chapter 2, haplogroup frequencies for mtDNA and the Y-chromosome combined with data on haplotype sharing can provide information about a population's prehistory along the maternal and the paternal lines respectively. Although genetic data cannot remove all ambiguities either, they do provide a more objective basis for the reconstruction of a people's prehistory and allow us to formulate hypotheses about shared ancestry and patterns of migration with more certainty. Recognising the fact that it is the contact situation as a whole, including socio-historical, ethnographic, linguistic and genetic data that must be considered for the most meaningful interpretation of the language data, this discussion will explore the data per contact setting rather than per linguistic domain.

After a schematic summary of the main lines of thought that were developed in the individual chapters, the contact situation between Dolgans and Evenks, and between Dolgans and Russians, are discussed separately. For both settings, I will summarise the contact-induced linguistic changes in lexicon, morphology and

syntax, and I will interpret them within the socio-historical context (such as status of the languages in contact, the attitudes of the communities in contact, the sizes of the communities, etc.) and link them to the genetic data.

9.2 SUMMARY OF RESULTS

Table 9.1 presents a schematic overview of the contact-induced changes in Dolgan. They are sorted by contact language on the one hand and linguistic domain on the other, as well as by the underlying processes of transfer. In addition to the two concrete contact languages Evenki and Russian, lingua franca is added as an explanation for those changes that probably arose as a result of the ancestor language of Dolgan as a means of intergroup communication. Crucially the explanations are not mutually exclusive, and changes can occur in more than one column, allowing for multicausality. This summary is a brief reminder of the investigated contact-induced changes and their associated processes and is meant as a mnemonic during the detailed discussion of each contact situation.

Table 9.1. Summary of contact-induced changes in lexical, morphological and syntactic domains

	EVENKI		RUSSIAN		LINGUA FRANCA
	Borrowing	Imposition	Borrowing	Imposition	
Lex.	• cultural terms • non-cultural terms	• semantic structure of kinship terms	• cultural terms • non-cultural terms • conjunctions		
Morph.		• regularisation - verb <i>e-</i> - unstable noun stems			• regularisation - verb <i>e-</i> - unstable noun stems
Synt.		• increased frequency of habitual use of <i>onton</i>	• word order • coordination strategies - subordination strategies	• word order • coordination strategies - subordination strategies	• parataxis • morph.synt. simplification

9.3 CONTACT WITH EVENKS

9.3.1 LINGUISTIC CHANGES

This research has shown that the contact with the Evenks has not just led to changes in the culture of the Turkic people who moved to the north, but has also had a linguistic impact and accounts for a significant subset of the differences between Dolgan and Sakha that we witness today. Most directly, this influence can be seen in full copies of lexical items for cultural as well as some non-cultural items, where form and meaning are copied wholesale from Evenki into Dolgan. More indirectly it can be seen in structural changes, such as the semantic restructuring of kinship terminology, the frequent use of the habitual mood in Dolgan, and the morphosyntactic properties and frequency of the coordinating element *onton* ‘and then’. In addition it was argued that the regularisation of the inflectional paradigms for *e*- ‘to be’ and of unstable noun stems could be attributed to a large number of L2 learners and the function of Dolgan as a language of intergroup communication. Based on the available socio-historical information on the relations between the native populations of the Taimyr Peninsula between the 17th and 20th centuries, and on insights from language contact theory, complemented by a certain amount of common sense, the identified changes were associated with the linguistic processes of borrowing and imposition (Van Coetsem 1995, 2000). While the changes cover a variety of linguistic domains, they have one characteristic in common: apart from the lexical copies that most probably were introduced into Dolgan through the process of borrowing, all other differences are structural changes associated primarily with the process of imposition. This of course does not exclude the possibility that at speaker level some of the changes developed as a result of more than one process: depending on individual differences in the language dominance of the speakers, the same linguistic variant could be brought about through the processes of borrowing and imposition in different individuals. However, the conclusions about the paramount linguistic process are based on generalisations that emerge from social and historical facts, which allow us to formulate hypotheses about the linguistic balance in the majority of bilingual at community level.

As described in Chapter 2, the ancestor language of Dolgan (referred to as Dolgan/Sakha) was the lingua franca on the Taimyr Peninsula in the 18th and 19th centuries and was used for intergroup communication for about two hundred years. This social dominance of Dolgan/Sakha in this region is a compelling

reason to assume that in this contact situation the non-Turkic-speaking populations would learn Dolgan/Sakha, rather than the other way round. This would have produced a considerable number of L2 speakers of Dolgan/Sakha, which is in turn associated with structural changes and the process of imposition (see Sections 3.1.3.3 and 3.1.4.1).

Lexicon

Evenki influence on the Dolgan lexicon materialises as full lexical copies, and as copies of semantic structure. Lexical copies from Evenki are not restricted to particular semantic fields, but are distributed across the entire lexicon. The overall number of lexical copies is not very high (22.5% of all lexical replacements and only 3.7% of all lexical differences between Dolgan and Sakha), but their distribution across a wide range of semantic fields points to a contact situation that went beyond the adoption of only cultural features and was not confined to the transition to a lifestyle of reindeer herding (even though many copies are indeed related to these practices).

The semantic changes that took place in the semantic fields of ‘the body’ and ‘kinship’ lend more credence to this idea, since all but one of the changes in terms related to ‘the body’ were such that the modified meaning matches the semantic pattern of Evenki. Despite this striking similarity, a language-internal explanation could not be excluded completely, since most semantic changes followed pathways that are common in language-internal change as well. Quite possibly these are instances of language change where multiple motivations conspired towards one linguistic outcome.

In contrast, the match in the semantic structure of kinship terms in Dolgan and Evenki is too striking and too particular to be caused by language-internal change. First, the matches in meaning are exact, and second, the change is not restricted to independent lexical items, but applies to an entire set of interrelated concepts, revealing that the entire system of kinship terms is affected. To recapitulate, it was found that Dolgan speakers label the concepts BROTHER/SISTER, UNCLE/AUNT and FATHER-IN-LAW/MOTHER-IN-LAW with lexical forms that match those in Sakha, but their meaning is allocated according to the Evenki system of kinship terms. The nature of the changes themselves, as well as the socio-historical information on the relation between the Dolgans and the Evenks between the 17th and 19th centuries (see Chapter 2) unite towards an

explanation of this phenomenon in terms of imposition of semantic structures from Evenki onto lexical forms of Sakha origin.

First, structural change, including changes in semantic structure, is typically associated with imposition, a correlation made on the basis of cognitive principles of L2 learning (see Section 3.1.3. for details) and confirmed by data from research on contact-induced change. However, there are more, and perhaps more compelling, reasons to arrive at the conclusion that the Evenki-speaking population, and not the speakers of Dolgan/Sakha, initiated these changes and projected the semantic structure of their mother tongue onto their L2 (Dolgan/Sakha). In addition to the fact that structural features of a speaker's L1 show through most persistently in his L2 as a result of cognitive learning principles, the nature of the semantic domain in which these particular structural changes took place also favours a scenario of imposition. Kinship terminology is a semantic domain for which there is a particularly tight connection between linguistic labelling and actual social structure. Kinship terms are not simple denotations of individuals, but they reflect the underlying social system of family relationships within a community. In that case, a scenario of borrowing becomes highly unlikely. It would mean that the combination of Evenki semantic (and social) structure with Sakha terms that we observe in Dolgan came about through L1 (and dominant) Sakha speakers who adopted the Evenki social, and consequently semantic, structure through borrowing, and matched this new structure onto their native terminology. While the adoption of a different social system would perhaps be possible in a situation of intense long-term cultural contact and high social and cultural pressure, it would be implausible that this happened without extensive borrowing of linguistic substance in this semantic domain.

More realistic is a scenario in which groups of Evenks joined the community along the Khatanga Trading Way, bringing with them their own customs, traditions and social structures. Dolgikh's table of marriages (see Section 2.3.2.3) showed that a considerable number of Evenks intermarried with the Turkic speaking Dolgan, and probably adopted Dolgan/Sakha as their L2. Other Evenks acquired Dolgan/Sakha as an L2 because of its use as a *lingua franca*. This setting supported a constantly renewed stream of L2 learners, who, importantly, did not (have to) abandon their own culture, but only used Dolgan/Sakha forms to label their own concepts. Some of these Evenk groups eventually shifted to Dolgan/Sakha due to the wide functional domain of this language, as well as its

perceived prestige. To apply this to the semantic change of kinship terms, one may assume that despite the linguistic adaptation to the Sakha, these Evenks did not adapt their own social structures during this process. They maintained their traditional social structure, but had to use Sakha terms to express these relations. This meant that through interlingual identification they had to find a match between their native Evenki terms and their closest equivalent in Sakha. In certain cases this match between the Evenki and the Sakha terms was perfect, but in other cases the semantic overlap was only partial, leading to differences in denotation of the Sakha term between L1 speakers and L2 speakers. Assuming a relatively large number of mixed Sakha-Evenk families, as well as of L2 Dolgan/Sakha-speaking Evenks, this 'foreigner' version of Dolgan/Sakha gradually became common use among the L1 speakers of Sakha as well, thus becoming the established way of using kinship terminology within this variety.

However, the most compelling reason to attribute this semantic change to imposition is the genetic confirmation of the intense contact between Dolgans and Evenks. The high proportion of Tungusic related haplogroups in today's Dolgan population indicates that an influx of Tungusic genetic material into the Dolgan community was certainly not exceptional and thus that marriages between Evenks and Dolgan/Sakha people were common. The fact that the Dolgans speak a Turkic language today implies that many Tungusic speakers must have gone through a stage of learning Dolgan/Sakha as a second language, endorsing the idea of imposition as an explanation for the semantic change. A similar scenario can also be seen to account for other contact-induced changes in Dolgan.

Regularisation

The discussion of regularisation patterns in the paradigm of *e*- 'to be' as well as in the inflectional paradigm of unstable noun stems concluded with the remark that it is not possible to make a rigid distinction between processes of L1 and L2 acquisition as an explanation of this development. Regularisation of paradigms is a common phenomenon in both domains, as a result of general mechanisms of human learning, and unless there are languages without any L2 learners the two processes cannot be rigorously separated. However, there are reasons to assume that in this particular case the explanation for the observed regularisation requires a component of L2 learning, and hence of contact. The main reason for this assumption is the geographical distribution of the regularised variety. Although not all existing Sakha dialects could be sampled for comparison, only

Dolgan, which is spoken in an area of contact (and which linguistically appears equally close to Sakha as Sakha dialects are to each other), displays this tendency towards regularisation.

One may object that the Dolgan-speaking area is not the only area where Sakha-speaking communities were in frequent contact with other ethnic groups, and that there are many examples of Sakha groups maintaining close relations with Tungusic-speaking groups (Evenks, Evens) where this regularisation did not occur. However, in these situations there is no genetic evidence of admixture between Sakha and Evenks, showing that these contact situations were fundamentally different from the one on the Taimyr Peninsula (Pakendorf pers. comm.). In addition, as rightly pointed out by Thomason (2010) one should not aspire to explain why changes did *not* occur in some communities, while they did occur in others. Rather one should aim at explaining the changes that *have* occurred in order to get better insights into the range of possible contact scenarios and their linguistic outcomes. For these reasons, the absence of regularisation in contact situations of Dolgan's closest relative Sakha does not at all demote L2 learning as a relevant explanation for the present change in Dolgan. Thus, the argument of geographical distribution and the clear genetic evidence of contact between Dolgans and Evenks are supported by the overall picture of differences between Dolgan and Sakha which is now emerging, and which reveals more examples of structural changes typically associated with L2 learning.

Habitual

One of these changes is the significant difference in the use of the habitual participle in Dolgan and Sakha. Frequency analyses of text corpora of Dolgan, Sakha and Even, as well as initial data from three Evenki dialects and Udighe, showed an interesting pattern: languages belonging to the same language family displayed significant frequency differences in their use of the habitual participle, while this difference disappeared between languages belonging to different families, but spoken in adjacent areas. In other words, there were significant differences between Dolgan and Sakha, and between the Evenki dialects, but not between Dolgan and the Ilimpijskij dialect of Evenki, which is spoken on the edge of the Dolgan-speaking area. While this could plausibly be argued to point to an areal phenomenon, it does not give an indication as to which languages behave 'typically' for the language family, and which languages have changed. Comparison with Tungusic Udighe and Even showed that the habitual in these

languages is used even more frequently than in the Evenki dialects, which would suggest that the change occurred in Dolgan, rendering it more similar to the Tungusic pattern.

Research on frequency comparisons using text-based corpora in studies of language contact is still in its infancy and more corpus data are needed to verify these hypotheses in the future. However, these preliminary results, in combination with the other structural changes and the socio-historical information about the area, support the conclusion that this may be one more structural change motivated by contact.

*Coordinative element *onton**

This social setting also provides the perfect conditions for the development of the final set of differences between Dolgan and Sakha associated with Evenki influence: the frequent use of *onton* ‘and then’ to link coordinate sentences and the potential role of this element in the absence of *uonna* ‘and’. In Chapter 8 two explanations were suggested for this overwhelming use of *onton*. The first was entirely language-internal and relied on the functional overlap between *onton* and *uonna* that could have rendered one of the two elements redundant, and could have eventually led to the loss of *uonna* in Dolgan discourse. However, this would not explain the complete absence of this element in the Dolgan language material available to me. The alternative explanation for this difference suggested influence from Evenki in the same way as described for the semantic restructuring of kinship terms, the difference being that this is not a content word but a function word. In a similar fashion to the semantic restructuring of kinship terms, it was suggested that L1 speakers of Evenki projected the semantic and functional properties of the Evenki coordinator *taduk* ‘and then’ onto the Dolgan/Sakha element *onton* ‘and then’ through interlingual identification, which was stimulated by the identical morphological structure of the two elements and their partial functional overlap. This would explain the difference in use of *onton* between modern Dolgan and Sakha (inter- as well as intraclausal use in Dolgan as opposed to merely interclausal coordination in Sakha), as well as the complete absence of *uonna* in Dolgan: rather than assuming that it gradually lost territory, in which case some kind of record, relic in fixed expressions or perhaps recognition by Dolgan people would be expected, it may simply never have been present in the L2 variety of Dolgan/Sakha that came to dominate the area.

9.3.1.1 A NOTE ON PHONETICS

Due to fundamentally different methodological requirements, an in-depth investigation of phonological, phonetic and intonational differences between Dolgan and Sakha was not included in the current study. However, since theories of language contact predict that changes in these domains are typically twinned with changes in syntax (Tomason and Kaufman 1988: 50) a brief overview of the literature on this topic, complemented by my own observations, is in order. This brief description deals with the allophonic variation of [s] and [h], [k] and [q] as well as [g] and [ɣ], and with the less strict adherence to the rules of vowel harmony in Dolgan. Since here the focus is on the phonetic realisation of the phonemes, IPA symbols are used for their representation. The correspondence between the transcription symbols used elsewhere in this thesis and the IPA symbols can be found in the reference information. To understand how the phonetic differences fit into the sound system, the consonant and vowel inventories for Dolgan and Sakha are presented in Tables 9.2 and 9.3. The square brackets indicate that this sound has an allophone.

Table 9.2: Consonant inventory of Dolgan and Sakha

	Labial	Alveodental	Palatal	Velar	Uvular	Glottal
plosive	b, p	d, t	dʲ	g, k	[q]	
fricative		s			[ɣ]	h
affricate			[dʒ], tʃ			
nasal	m	n	ɲ	ŋ		
glide			j			
liquid		r, l				

Table 9.3: Vowel inventory in Dolgan and Sakha

	Monophthongs				Diphthongs	
	low		high		unrounded	rounded
	Unrounded	rounded	unrounded	rounded		
back	a, a:	o, o:	u, u:	u, u:	ua	uo
front	e, e:	ø, ø:	i, i:	y, y:	ie	yø

Allophones [s] and [h]: allophonic variation between [s] and [h] is common in both Dolgan and Sakha, as in *sa:s* and *ha:s* ‘spring’. Since the variant with [s] is represented in Sakha orthography, and the variant with [h] in Dolgan

orthography, this seems an established difference between the languages. However, in fact it is confined to the orthographic domain, since in spoken language Sakha and Dolgan speakers mostly use [h]. Nonetheless, there could be a difference in the explanation of the use of [h]. The replacement of [s] by [h] has been ascribed to substrate influence from Evenki in the literature for both languages, since the distribution of the allophones matches the distribution of the same allophones in the Evenki dialects they were in contact with (Ubryatova 1985: 32). However, on the basis of historical word lists of Sakha, in combination with historical and genetic data, Pakendorf argues that for most dialects of Sakha a language internal motivation is a more likely explanation (Pakendorf 2007: 93). However, for Dolgan she concludes that an external explanation cannot be excluded.

Allophones [k] and [q]: These velar and uvular sounds also occur as allophones in both Dolgan and Sakha. However there are differences with respect to the details of their phonetic realisation and their distribution. In Dolgan [k] and [q] are both plosives, whereby [k] is velar and [q] is uvular (Stachowski 1999: 17). In Sakha [q] is more aspirated and is therefore sometimes classified as a uvular fricative [χ] (Stachowski and Menz 1998: 418). According to my own observation, the realisation of this uvular sound in Sakha varies across speakers, some producing a more plosive, others a more fricative uvular. As far as the distribution is concerned, in both languages [q] occurs before and after low back vowels (/o/ and /a/), as in *χa:r* 'snow' and *hoyotoχ* 'single'. However, in Sakha this allophone is also used after low front vowels, whereas in Dolgan [k] is used in this environment (*belex* vs. *belek* 'present'). Ubryatova notes that the Dolgan distribution becomes more frequent also in the northwestern dialects of Sakha in a similar fashion to the difference in variation between [s] and [h].

As a motivation for this difference it is worth noting that the allophone [q] is absent in Evenki. The correlation between decreased use of this allophone in Dolgan and the northwestern dialects of Sakha, and increased intensity of contact as we know it from historical records suggests that substrate influence from Evenki should be considered as an explanation for this difference in distribution.

[g] and [ɣ]: Like its voiceless counterpart, the voiced velar plosive /g/ also has two allophones: a uvular voiced approximant [ɣ] and a voiced velar plosive [g]. Both in Dolgan and Sakha, [ɣ] is used between low back vowels as in *ɑya* 'father' and *oyo* 'child'. However, in Sakha it is also found before and after low vowels and between low front vowels, whereas in Dolgan, [g] is used in these environments

(cf. Sakha *beyehe* and Dolgan *begehe* ‘yesterday’). According to Ubryatova, the limited use of [ɤ] in Dolgan could be due to influence from Evenki, in which this allophone is used in the same phonological environment as in Dolgan (Bulatova and Grenoble 1999: 5).

The final difference concerns the rules of vowel harmony. In principle, both languages apply rules of vowel harmony on two dimensions: a) back vs. front; and b) rounded vs. unrounded. This means that words contain either only back or only front vowels (resp. *balik* ‘fish’, *ijedes* ‘face’) and that these vowels are either all rounded or unrounded (resp. *törüt* ‘ancestor’, *ijedes* ‘face’). This applies to word roots, as well as between roots and suffixes. The only exception to this rule are high rounded vowels (/u/ and /y/), which are followed by unrounded vowels when they are low (i.e. /a/ and /e/, instead of /o/ and /ø/). For example, rounded *munnu* ‘nose’ gets the unrounded suffix *-ta* in *munnu-ta* [nose-POSS.3SG] ‘his nose’ and not **munnu-to* as would be expected in a consistent system of labial vowel harmony.

While this rule applies almost without exception in Sakha, Dolgan allows for more variation. Within lexical roots, inconsistencies are mostly found in copies from Evenki (e.g. *bugdi* ‘spotted’ *gedalun* ‘dragon fly’) or from Russian (e.g. *abiet* ‘lunch’, *ha:sturuga* ‘snow groove’), but also Turkic words with compound etymology in some cases do not conform to the system of vowel harmony, e.g. *harsierda* ‘morning’, which contains the elements *harsin* ‘tomorrow’ and *erde* ‘early’ is in Dolgan pronounced with a front diphthong, whereas in Sakha it is pronounced [sarsuarda] or [harsuarda], following the rules of back-front vowel harmony. Across morphological boundaries, primarily non-native lexical items are affected, as can be seen from the comparison of Dolgan and Sakha inflection of the word *hiliiep/kiliep* (> Russ. *xleb*) ‘bread’, which is inflected with a back vowel in Dolgan but with a front vowel in Sakha.

DOLGAN

- (9.1) *min dnevnip-par huruj-uom*
 1SG diary.R-DAT.1SG write-FUT.1SG
 ‘I will write in my diary.’ (IMA: sound file)

SAKHA:

- (9.2) *oloχ-χo ti-edd-en kilieip-ten kilieip-ke ti-edd-en*
 life-DAT reach -CAUS-SQ.CV bread.R-ABL bread.R-DAT reach -CAUS-SQ.CV
 ‘You live from bread to bread.’ (REX: 248)

The looser rules of vowel harmony in Dolgan have been described by other Dolgan specialists (e.g. Artemyev 2001: 49, Ubryatova 1985: 21), and Ubryatova goes so far as to say that in Dolgan “the law of vowel harmony... is no longer an obligatory regularity” (Ubryatova 1985: 21). While this is in my opinion an exaggeration, considering the productive application of vowel harmony in native words in Dolgan, my Dolgan language material shows that the situation is different for non-native elements since back-front vowel harmony barely applies to these lexical items.

While the current intense contact with Russian (and consequent attrition) may seem an obvious explanation for this difference, it loses pertinence when we recall that Ubryatova’s data were collected in the 1930’s. At this time, Dolgan could still be convincingly called the dominant language on the Taimyr and it is unlikely that rules of the Russian sound system would have affected Dolgan at this stage. Alternatively, substrate influence from Evenki could be part of the explanation. In this language, vowel harmony is not conditioned by a distinction between front and back vowels as in Dolgan and Sakha, but by a distinction between high and low vowels instead (Bulatova and Grenoble 1999: 4). Mid and low vowels, except neutral schwa (i.e. /e:/, /a/, /a:/ /o/, /o:/), combine with suffixes containing the vowel /a/. High vowels (/i/ and /u/ and their long varieties) do not conform to vowel harmony and combine with suffixes containing the neutral /ə/ or the low vowel /a/. The choice between these two variants is conditioned by a historical merger and is not transparent from a synchronic point of view (Bulatova and Grenoble 1999: 4).

Now while learning Sakha as their L2, it is possible that Evenks quickly picked up on the distinction between low and high vowels, since these are meaningful categories in Evenki as well, whereas they may have paid less attention to the division between front and back vowels which is so important in Turkic. The widespread use of the low vowel in Evenki suffixes (as opposed to neutral /ə/, and high vowels) may have rendered the low variant the default form. As a result, this form and its combinatorial properties may have been projected onto this L2 variety of Sakha: it became used with back and front vowels, which is what we see in the inflection of the non-native lexical items in modern Dolgan.

The question remains why this less stringent application of vowel harmony applies to foreign items in particular. While this needs to be investigated in depth, a possible explanation is that initially the inflected forms of these foreign items did not occur in the native language input. In contrast to native items, their

inflected forms were not stored as a single phonological unit in the brain, and inflectional rules had to be applied productively. Since the roots of some of these foreign items have an unusual form with respect to rules of vowel harmony as well, it is likely that in such cases the default (i.e. low vowel) form of the suffix was taken as the default solution.

9.3.2 INTERPRETATION OF THE LINGUISTIC DATA WITHIN LANGUAGE CONTACT THEORY AND GENETICS

9.3.2.1 CONTACT BEFORE ARRIVAL ON THE TAIMYR

The genetic profile of the current Dolgan population shows a varied pattern. As was discussed in Chapter 2 they share a large amount of their mtDNA with other populations, in particular with the Taimyr Evenks and the Yakut-speaking Evenks of the Olenek district (see Section 2.6.2 for more details). It was shown that the measure for population difference (the F_{st} value) is so low that these populations can be said to share a single genepool in the maternal line. While mtDNA sequences are shared all across Siberia, the Dolgans, Taimyr Evenks and Yakut-speaking Evenks share more than any other pair of populations in Siberia. Thus, while we cannot draw any definite conclusions concerning the interactions between specific pairs of populations, the extremely high frequency of sharing between these groups on the Taimyr Peninsula and its neighbouring regions does indicate a considerable amount of gene flow in the maternal line, which could be due to recent common ancestry, to intermarriage or to both.

The Y-chromosome reveals a more distinctive pattern. Compared to neighbouring populations, the haplogroup complement of the Dolgans is more differentiated, and includes haplogroups that appear to come from different sources in roughly equal proportions. The STR analysis of haplotypes showed that haplogroup N3 was shared with the Sakha, and haplogroup C has its origins in the Evenk population. For haplogroup N2, the third main component of the Dolgan genetic material, the origins could not be established unambiguously since identical STR sequences were found in comparable numbers of Samoyedic and Tungusic individuals. Accepted reasoning is that if the sharing of haplogroups goes back to a common ancestor, it is unlikely that individuals today will still share the exact same STR haplotype on the Y-chromosome, whereas this is plausible in a scenario of more recent contact. Therefore it is most likely that the mixed pattern

in the paternal line is the result of more recent admixture. Thus, the genetic results provide some crucial insights into the history of the Dolgans, including the rough time period of their formation as a separate ethnic group.

Contact settings between Tungusic and Turkic groups become relevant for the formation of the Dolgan people from the moment Evenk and Sakha clans began to populate the area around the Lena and Vilyuy rivers. While we know that the first Sakha people arrived at the Lena River in the 13th century (see Section 2.3.2.2) we do not know exactly when the contact setting with the Tungusic clans began to take shape. However, it was an established situation by 1638 when the Russian officials registered both Turkic Sakha and Tungusic Dolgan clans in this region. Importantly, at that time both groups recognised the same Sakha headman. While the fact that two ethnolinguistically different groups were ruled by only one headman is interesting in itself, this fact must be attributed additional significance in the light of the presumed Evenki-Sakha bilingualism in the Dolgan people.

Dolgikh, who traced back as many of the various ‘components’ of today’s Dolgan population as he could, reaches the same conclusion in his detective-like work ‘The origin of the Dolgans’. He confirms that the Tungusic Dolgan clan, which constitutes an important proportion of today’s Dolgan population, inhabited the area around the Lena and Vilyuy Rivers in the 17th century. If this clan, and the Dolgan clan mentioned by Ubryatova (see Section 2.4.1) are the same, then we can infer that this Tungusic Dolgan clan lived in the Lena and Vilyuy area in close vicinity to the Sakha people and acknowledged the Sakha headman.

Despite the fact that we know little about the exact nature of the relations between different indigenous groups themselves, including their use of, and attitude towards, other languages, there are indications that the beginning of the 17th century may have been the seminal moment for incipient bilingualism within this Dolgan clan. If Dolgikh’s interpretation is correct, it was the Tungusic clans who adopted Sakha as a second language rather than the other way round. He attributes this skewed balance to numerical dominance of the Sakha in this ethnically mixed area of the Lena and Vilyuy as well as the Olenek basin. He says that,

We also know that towards the end of the 17th century on the Olenek River the Yakuts might have even outnumbered the different Evenk clans. Therefore it is

very probable, that among the Dolgans knowledge of the Sakha language was widespread.¹

The fact that Dolgikh sees a causal relation between the large number of Sakha people and bilingualism in the Evenk community illustrates that these two ethnic groups were not indifferent towards each other, and that their presence in the same area was not restricted to pure coexistence. He makes this explicit when he explains a sudden increase in members of the Dolgan clan between 1678 and 1761 by the possibility that Sakha people merged with the Tungusic Dolgan population (Dolgikh 1963: 110). What exactly he has in mind when he talks about ‘merging’ is unclear. He does not specify his ideas as to whether the Sakha settled among the Tungusic Dolgans, or that there was also frequent intermarriage between the two groups. However, either scenario of admixture would result in the shared haplotypes in the mtDNA as well as the Y-chromosome that we see in the Dolgan population today. If it is true that the Tungusic Dolgan clan was ruled by a Sakha headman in these years, it is possible that the Sakha had not only a numerically, but also a socially dominant position, which in turn would be further justification for the more powerful position of Sakha in the community.

While there were obviously individual bilingual speakers among the Sakha as well, the more widespread bilingualism in the group of Dolgans, motivated by the numerical and social dominance of the Sakha, would have had primarily a linguistic impact on the version of Sakha spoken by the Dolgans as a second language. Since in all likelihood these Dolgans were initially all dominant in Evenki, the linguistic consequence that presumably appeared first is structural variation in the target language (Sakha) due to imposition. This is perfectly compatible with the Evenki-induced changes that we see in modern Dolgan. The characteristic use of the habitual participle, the use of *onton* in clause combining, as well as regularisation and reanalysis could well have their origins in this social setting. Also the changes in the semantic structure of kinship terms are sensibly explained within this configuration. However, since contact between Turkic and Tungusic-speaking groups became much more intense once they had migrated to the Taimyr Peninsula, it is most likely that the majority of these changes became established in this contact variety of Sakha from the 18th century onwards.

¹ Мы знаем также, что к концу XVII в. на Оленеке якутов было едва ли не больше, чем эвенков разных родов. Поэтому вполне вероятно, что знание якутского языка у долган было широко распространено. (Dolgikh 1963: 110, translation mine).

After this initial period of contact between Dolgans and Sakha in the Vilyuy and Lena basins during the first decades of the 17th century, during which some initial bilingualism may have developed, we know that these groups retreated further to the north. Since one of their purposes was to dodge the Russian tax collectors, it is no surprise that we have no explicit information about how they lived through these years, and how much contact there was with other groups. After all, it was the tax collectors who provided what sparse information we have on the indigenous populations of Siberia in the 17th century. However, if the situation reconstructed above is correct, and the group of Dolgans that arrived on the Taimyr Peninsula in the late 17th century were the same people (and their offspring) as the potentially bilingual group of Dolgans, one can be confident that the linguistic variation that was initiated during the first period of contact was maintained throughout these years. This inference is based on the assumption that the years of ‘retreat’ to the upper regions of the Vilyuy and Olenek Rivers which occurred between 1655 and 1678 (Dolgikh 1963: 108), were spent in relative isolation, removed from large groups of L1 Sakha speakers that would have levelled out the Evenki influence in the Dolgans’ variety of Sakha.

9.3.2.2 CONTACT ON THE TAIMYR

Contact with other ethnic groups resumed after arrival on the Taimyr Peninsula, and intensified after the appearance of the Russians. The arrival of Russian explorers, merchants, and their accompanying personnel in arctic Siberia generated not only new contact settings between themselves and the native Siberian people, but also enhanced the contact between indigenous peoples. The main catalyst for this increase in interethnic socialisation was the Khatanga Trading Way. This corridor of permanent trading stations traversed the Peninsula from west to east and connected places as well as people in a more conspicuous way than the nomadic routes of the Tungusic and Samoyedic people had done so far. The trading activities acted as a magnet for people from different ethnic groups.

As we know, the Dolgans became the main protagonists in this way of life. While today the term refers to a clearly defined and seemingly homogeneous ethnolinguistic group, Dolgikh’s analysis shows that their ethnic origins are convoluted and comprise more ethnic groups than the Tungusic Dolgan clan and

groups of Sakha that came from the Lena and Vilyuy. His investigation shows that at least the Tungusic Dongot, Edyan and Karanto clans, Turkic Sakha from the Tundra (Olenek region) as well as Russian tundra peasants make up a substantial part of today's Dolgan population. This means that the initial contact between the Dolgan clan and Sakha in more eastern regions was complemented by cultural and linguistic influence from other groups.

Armed with this knowledge, we may ask the question why of all these different groups it was the name of the Dolgan clan unified with the Sakha language that became the main markers of the people leading this life of reindeer herding and trade. Since we just concluded that a clan with both characteristics potentially existed before arriving on the Taimyr, the idea that the establishment of their name signifies their leading position is tempting. However, the history of the Dolgan people shows clearly that we must take extreme care in relying on labels alone. Nevertheless, even if today the name 'Dolgan' denotes a population with a wider range of ethnic origins than just one Tungusic clan, the fact that this name was chosen to represent a large proportion of the Taimyr's native population whereas the other clan names have fallen almost into oblivion (e.g. Dongot, Karanto), may carry historical significance that goes beyond mere political decision-making.

One possible explanation is that the Khatanga Trading Way and the Dolgan clan appeared almost simultaneously on the Taimyr Peninsula. Disconnected from their homeland, one could imagine that their position as 'newcomers', who were not yet as established in this territory, as were some of the other Tungusic and the Samoyedic groups, made them more open and keen to engage in a new lifestyle and occupy this new niche. While people with other ethnic backgrounds surely took part in these activities as well, their established routine and traditional ways may explain why they identified with this way of life to a lesser extent.

The situation with the Sakha is more complicated. First, the Sakha from the Olenek area were, like the Dolgans, newcomers to this region, and had arrived in the basins of the Kheta and Khatanga by the end of the 17th century (Dolgikh 1963: 117). In fact, they were one of the first 'immigrants' in this area who had headed this way because of a famine in 1681 - 1682 resulting from a change in migration routes of the wild reindeer in the Olenek area. Driven by hunger, these Sakha groups set off with their dog sleighs in the direction of the Kheta and Khatanga rivers in the hope of finding food. On arrival they probably met the abovementioned Tungusic Dolgan and other Sakha groups who came from the

Lena and Vilyuy area. So if the above reasoning makes sense, the Sakha would have been equally suited to engage actively in, and become associated with, this position. Second, since the modern Dolgans are linguistically so closely related to the Sakha, how can we be sure that it was not in fact the Sakha who engaged mainly in the trade, and adopted a Tungusic name, rather than a Tungusic group that had shifted language?

While trying to solve this question, the confusing history of nomenclature in the Dolgan people described in Chapter 2 does not exactly help matters. However, the genetic profile of the current Dolgan population and the linguistic characteristics that set them apart from the Sakha provide more reliable cues, and from their combination crystallises a picture that allows space for both events. The following paragraphs will elucidate how.

If the present Dolgan population were predominantly Sakha who adopted, or were given, a Tungusic name, we would expect their genetic material to consist primarily of haplogroups associated with the Sakha population. Since traditional Sakha communities are patrilocal, this would mean that a high proportion of haplogroup N3 would be expected. With respect to the mtDNA we would not expect any clear patterns, since exogamy is widespread, and women would have come from different ethnic groups. While the mtDNA of the Dolgans is mixed as expected², the Y-chromosomal haplogroup distribution does not evidently group the Dolgans with the Sakha. In fact, the picture looks highly mixed. While the Dolgans share about 40% of their genetic profile in the paternal line with the Sakha, nearly 30% and up to 49% is potentially of Tungusic origin, depending on whether haplogroup N2 in the Dolgans (which is associated with Tungusic and Samoyedic populations) has Tungusic or Samoyedic origins. Thus, to say that the Dolgans are Sakha people who were given a different name would not account for the amount of Tungusic admixture that is clearly represented in their genetic profile. The same issue arises when we classify them as Tungusic, because this leaves the 40% of Sakha-related genetic material unexplained.

The linguistic picture leads to the same conclusion. While Evenki influence is not overwhelming, we find lexical copies, as well as structural changes in lexicon, morphology, syntax and potentially phonetics. While the lexical copies may have

² And even more than expected, since the Dolgans share an extremely high amount of mtDNA sequences with the Taimyr Evenks and Yakut-speaking Evenks. The fact that there is more sharing between these groups than there is between Dolgans and Sakha indicates that Evenki women were incorporated into the Dolgan population.

been introduced by dominant Sakha-speakers through the process of borrowing, the structural changes were most probably initiated by Evenks who acquired Sakha as their second language. L2-speakers of Sakha. Most information about the social setting in which this may have occurred can be inferred from the changes in semantic structure of kinship terminology.

Here, one can imagine that during the acquisition process, the Tungusic Dolgans quickly acquired the Sakha lexical forms for kinship relations that matched kinship categories in their own social system. However, in cases where the semantic match between the Evenki and the Sakha word was not perfect, these distinctions may not have been picked up so easily, particularly when distinctions in the L2 system are more fine-grained than those in the L1. In such cases the semantic structure of the L1 (Evenki) was projected onto the lexical forms of the L2 (Sakha). It is worth mentioning that for all but one of the semantic changes in the kinship terms it is the terms used from a male perspective that have become established in today's Dolgan. For example, the Sakha terms for older sister were *edzij* from a male perspective and *ayas* from the perspective of a woman. In modern Dolgan, *edzij* has the meaning of older sister of both a man and a woman, indicating that only the lexical item that was originally used for men was adopted into the Dolgan version of Sakha. While the absence of a distinction between the male and female perspectives itself can be attributed to Evenki since it does not make such a distinction and expresses both perspectives by a single term, there is no linguistic reason why *edzij* should be favoured over *ayas*. Rather an explanation in social terms is needed.

Any inferences must remain tentative and are proposed with extreme caution, but this tendency could have arisen in a situation where Evenki-speaking women married Sakha-speaking men rather than the other way round. Through hearing their Sakha-speaking husbands refer to family members, Evenki-speaking women would have adopted these terms to refer to their own relatives as well. Since these Evenki-dominant members of the Dolgan community were not used to a distinction between male and female perspectives in their own language, they may not have been on the lookout for these extra terms, and this may have resulted in the merger of the male and female perspectives, using the terms typically used by male speakers.

In the absence of numerically large Sakha-dominant groups, the bilingual members of the Dolgan community would have transmitted their Evenki-inspired version of Sakha to the next generation, and what was linguistic variation in the

first generation of bilinguals may have started to take root and become established as new linguistic conventions. The hypothesis that a large part of the bilingual population consisted of Dolgan women who married Sakha men would be compatible with this scenario. Since in this culture it is the women who mainly take care of the children and spend most time with them, they play a crucial part in the early language development of their children. If a significant proportion of the women spoke Sakha as a second language, then this language variety quite plausibly was transmitted to their children as well. Thus, the genetic, linguistic and historical information all indicate that Evenks, who shifted their language, but also Sakha, who shifted their ethnic identity, played an important role in the shaping of the ethnolinguistic group that carries the name Dolgan today.

Contact between Tungusic and Turkic groups in the past, as well as a certain degree of bilingualism in the Dolgan clan, may have enhanced contact between the Sakha and Dolgan once again, and enhanced the use of the Sakha language as a means of interethnic communication. Once they had established themselves along the Khatanga Trading Way, their prevalence in this niche would have stimulated other ethnic groups to conform to their norms and use Sakha as the language for trade and interethnic contact, including other Tungusic groups and Russians. This may have accelerated the emergence of second language speakers and explain how Sakha acquired its status as a *lingua franca*. Those who completely identified with this new socio-economic community of Sakha and Dolgans around the Khatanga trading way most probably merged with the prevailing population and would eventually shift to Sakha, with inherent linguistic consequences.

The genetic profile of the Dolgans confirms Dolgikh's analysis that it must have been relatively large numbers of Tungusic people who made this choice. This would have reinforced the contact-induced variation due to imposition that had been initiated during the Turkic-Tungusic contact in the mid 17th century, as well as introduced new variation. On the other hand, the considerable number of native Sakha speakers for whom this was a contact situation of language maintenance, would have modified their language by copying lexical forms from Evenki, with emphasis on, but not restricted to, unfamiliar concepts such as terms for reindeer herding, which was new to them.

While Russian tundra peasants also took part in the process of Dolganisation, their genetic and linguistic impact at this stage seems limited. The following section will show that the changes due to contact with Russians are the result of the recent Russian dominance in the Siberian arctic.

9.3.3 SUMMARY

The above discussion has illustrated the multifaceted character of the relationship between Evenks and Sakha, which are shown to be the two primary ancestors of modern Dolgans. For a correct understanding of their complex history with its large gaps in documentation, a division must be made between the contact setting during the initial period of contact before arrival on the Taimyr and the situation that obtained after they had reached the far north.

In the first setting (1638 and earlier) Tungusic Dolgans and Turkic Sakha lived on the Lena and Vilyuy rivers. Facilitated by the fact that the Sakha were numerically and socially dominant, part of the Sakha population incorporated members of the Tungusic Dolgan clan, which may have led to incipient, but probably limited, bilingualism in the Dolgan clan. After several decennia of relative isolation in the upper reaches of the Vilyuy River, this clan reached the Taimyr Peninsula by the end of the 17th century. The significant number of second language speakers that developed there would have been responsible for structural (and possibly phonetic) variation in their version of the Sakha language. During these years, bilingualism increased as a result of intermarriage and the role of Sakha as a *lingua franca*, and initial variation in the use of Sakha by L2 speakers may have become established as new linguistic conventions, leading to a characteristic version or dialect of Sakha that was to become Dolgan. At the same time, the first Russian colonisers had appeared on the Taimyr and were in need of transport and other services. Dislocated from their homeland and potentially in need of material goods, the Dolgans may have been more susceptible to the newly created 'jobs' along the Khatanga Trading Way than other indigenous populations, and so may have become the main representatives of this lifestyle. Since the Sakha from the Olenek region were in a comparable position, they may have joined the Dolgans, thus leading to a second encounter between Dolgans and Sakha. If we are correct in assuming that at least part of the Dolgan population had already some command of Sakha at that time, the choice for Sakha as an intergroup language would make perfect sense. Its status as a *lingua franca* in turn would have led to a further increase in the number of second language speakers, who left their traces in the structural change that the language underwent (e.g. kinship terms, regularisation, simplification of clause combining). Since most second language learners were probably Evenks (Edyan, Dongot, Karanto), it is no surprise to find increasing substrate influence as a result of imposition from Evenki. Russian

substrate influence is presumably negligible at this stage, for reasons to be discussed below.

9.4 CONTACT WITH RUSSIANS

The linguistic differences between Dolgan and Sakha that have developed, and are currently developing, as a result of contact with Russians do not serve so much to disentangle issues concerning Dolgan prehistory as they inform about on-going processes of language change. Although contact between Dolgans and Russians existed from the moment the Russians arrived on the Taimyr, it seems that the first contact in pre-Soviet times was too sporadic to lead to any of the structural variation on the Dolgan language as described in this thesis.

Since we know that the activities around the Khatanga Trading Way were predominantly ‘managed’ by Dolgans or Dolgan/Sakha-speaking people, we can be certain that contact between Russians and these groups commenced at this time. However, apart from certain lexical copies for unknown concepts such as ‘bread’ and ‘sugar’, items that were introduced by the Russians to win the goodwill of the native population and that later became important trade items, the intensity of contact and thus the level of bilingualism was too low during this time to have any significant linguistic impact. In addition, we have no evidence that the Russian language enjoyed any particularly high status at the time, thus excluding the possibility that its influence may have been extensive despite low intensity of contact. These hypotheses are confirmed by the genetic profile of the Dolgans. STR-analysis has shown that Russian admixture is unmistakable, and thus confirms Dolgikh’s idea that intermarriage between Russians and Dolgans took place, but its proportion in the Dolgan population is not large. The haplogroups of European origin account for less than 10% of the Dolgan Y-chromosomal genepool. For a correct representation of the contact situation between Dolgans and Russians before the establishment of the Soviet Union, it seems sensible to recognise a division within the group of early Russians on the Taimyr between temporary visitors, including mainly governmental officials, merchants and their personnel, and the permanent inhabitants also known as the tundra peasants. While the first group probably knew very little Dolgan and would for that reason be unlikely to leave any Russian traces in the Dolgan language, the second group almost completely merged with the Dolgans, and since we have no indication that

their numbers were particularly high, any significant influence from second language learning is implausible.

With respect to the presence of Russian Y-chromosomal haplogroups in the Dolgans, they probably appeared through marriages between tundra peasants and Dolgan women. While the temporary visitors may have also occasionally had physical contact with indigenous women, for obvious reasons we have no exact information to what extent this was common practice, which makes it hard to estimate the share of these Russian visitors in the current Y-chromosomal profile of the Dolgan population. This situation of relatively low-intensity contact between Dolgans and Russians changed dramatically in the 20th century with the establishment of the Soviet regime, when indigenous peoples were forced to give up their autonomy and become part of Russian society. Russian became an obligatory means of communication with anyone in powerful positions, which was a practical incentive to acquire this language. Within two or three generations, the balance of bilingualism in Dolgan society changed from predominantly monolingual and certainly dominant in Dolgan, through a stage of balanced bilingualism, to the current situation in which Russian is dominant for the majority of children and for some is even their only actively used language. It is during this time of intense contact between Dolgan(s) and Russian(s) that the Russian-induced changes emerged in Dolgan. The next section summarises the Russian-induced changes discussed in this thesis, after which they will be positioned within their socio-historical context.

9.4.1 LINGUISTIC CHANGES

Lexicon

Russian copies are widespread in Dolgan as well as in Sakha. However, since the investigation of Russian copies was based on the restricted set of lexical items from the Loanword Typology List, and since the knowledge of vocabulary varies considerably across individuals, it is hard to give a realistic estimate as to how Dolgan and Sakha differ with respect to the overall percentage of Russian copies in their lexicon (i.e. beyond the meanings in the Loanword Typology List. Nonetheless, we can say with confidence that they are not restricted to a particular semantic domain, since they include cultural as well as non-cultural items. As mentioned above, many Russian lexical items entered the Dolgan

language as a corollary of new concepts that the Russians first introduced to the indigenous people, but there are also cases where Russian lexical items are used for phenomena that are entirely disconnected from Russian presence, such as *kumar* 'mosquito' (Russian *komar*) and *namuluox* 'swamp' (Russian *navalok*).

A more useful way to look at the Russian copies is to categorise them according to the different social settings in which they were introduced into the Dolgan language. These social settings can be seen to correspond temporally to the pre-Soviet period, the Soviet period and the post-Soviet period. During the pre-Soviet period, contact between Dolgans and Russians was predominantly a relation of trade. Previously unknown objects and activities were introduced to the Dolgan people, and entered the life, as well as the language, of the indigenous population as complete units of form and meaning. At this stage, Dolgan/Sakha was the dominant language in the Dolgan community, and even beyond, as a *lingua franca*, so therefore these Russian lexical items are rightly classified as copies transferred into Dolgan through processes of borrowing. These early copies are characterised by phonological adaptation to the Dolgan/Sakha sound system, and often refer to foreign cultural items and activities.

In the Soviet period, many Dolgans had to become, and became, bilingual in Russian. The working environment of adults and obligatory boarding schools for children were all in Russian. Nonetheless, for many people Dolgan, which now was recognised as a separate language, still remained the dominant language, but the social and cultural values adhered to in their traditional way of life were gradually overruled by the ones approved of by Russian society. The lexical items introduced during this time period are therefore also the result of borrowing, but they cover a wider range of semantic fields. Since the level of bilingualism was increasing and people had a better knowledge of Russian, these words are less phonologically integrated than the copies from the early years of contact.

After the forced settlement and obligatory education at the end of the 1970's, the Russian language gained more and more territory. Children were in a Russian-speaking environment from very early on, often not seeing their Dolgan-speaking parents for months at a time. This has led to the current linguistic situation where most children are dominant in Russian, many even monolingual. These children still have some passive knowledge of Dolgan and can partly understand their parents and grandparents when they converse in Dolgan, but they cannot actively use the language. The only exceptions are the easternmost villages of Syndassko and Sopochnoe, where Dolgan is still the default language in everyday life for

adults as well as for children. However, with the exception of a few very old speakers, everyone is perfectly bilingual in Russian, in particular the children. Although the lexical changes from this period do not necessarily look different from the ones introduced during the Soviet period, except that they are even less phonologically integrated into the Dolgan sound system, the underlying process is different. Since the latest additions to the lexicon are introduced by a generation whose dominant language is Russian, the appropriate process to describe these changes is imposition and not borrowing.

Word order

A similar division between different time periods was proposed to explain the higher flexibility in word order, in particular the frequent occurrence of SVO sentences in Dolgan when compared to Sakha. Since we do not have pre-Soviet Dolgan texts that could prove that SVO sentences were less frequent before intensification of Russian contact, we have to be careful not to draw too definite conclusions about the time this variation began to occur. The fact that this word order is found in all age groups could lead to the impression that the increased use of SVO order has nothing to do with contact and is a purely language-internal development, based on the cognitive 'heaviness' principle that makes people want to move longer constituents towards the end of the sentence. However, the clear correlation between the occurrence of these constructions and the vicinity of Russian-speaking centres and intensity of contact shows that Russian influence at least reinforces, if not causes, this variation.

While age does not influence the frequency of use SVO construction in Dolgan speech, it does influence the identification of the linguistic process underlying this variation at the level of the bilingual individual. As for the lexical changes, it was argued that on the level of the individual, the variation in word order is also best explained in terms of multiple processes, because there are differences in linguistic dominance between different age groups. All recorded speakers grew up in a bilingual environment. However, the oldest generation (>70 years) is dominant in Dolgan, whereas the majority of the two youngest generations (< 40 years) is dominant in Russian. Therefore, the increased use of Russian SVO word order was proposed to be the result of borrowing in a situation of intense contact for speakers in the first age category, and of imposition due to dominance of Russian in the second, for whom Dolgan sometimes even appears as an L2.

The age group between 70 and 40 is hard to allocate to either of these categories, because of its high amount of internal variation. They grew up in the period between 1940 and 1970, which was the transition period from a society in which Dolgan was the dominant language to a society in which Russian took this role. Since this process did not progress at the same rate for every village and every individual, there were individual differences in linguistic dominance depending on where they grew up, their parents' attitude, as well as their own aspirations. Therefore I assume that within this group the same linguistic outcome (i.e. word order variation) must be explained as structural borrowing in the Dolgan-dominant Dolgans but as imposition in those whose dominant language had already become Russian. The development of this variation in general may have been facilitated by the fact that the change from SOV to SVO is also a commonly occurring language-internal change as well.

Clause combining

The changes in clause combining can be explained in a similar fashion. The introduction of Russian coordinating and subordinating conjunctions into the Dolgan language, and the syntactic consequence of rendering the position of the conjuncts in the sentence more flexible, were explained through a process of borrowing for the speakers who are dominant in Dolgan (roughly >70 years), whereas this change was attributed to imposition in the younger, Russian-dominant generation (<40 years). For the same reasons as were given for word order change above, the generation in-between shows too much individual variation to be classified into one of these categories in a sensible way. The same multicausal explanation was proposed for the pseudo-coordinate construction with *gitta* 'with', although the marginality of use of this construction, as well as its occurrence in the speech of people who use Russian every day, favour an explanation in terms of imposition.

The reduction of morphosyntactically complex constructions and the smaller selection of frequently used coordinators were associated with on-going language attrition and with the use of Sakha as a lingua franca (see Section 9.5 for details). Through the rapid spread of Russian and the gradual deactivation of Dolgan in bilingual speakers, certain details of the latter language are lost and not passed on to the next generation. This more constrained variety of Dolgan is then acquired and becomes established as the new norm in the community.

While simplification is a common process in language attrition and therefore a plausible explanation for the phenomena observed in Dolgan, the preference for paratactic syntactic constructions over morphosyntactically complex ones could also be due to the function of Dolgan as a lingua franca. Its function as an intergroup language would have favoured a more transparent structure and avoidance of morphosyntactically complex, irregular or idiosyncratic constructions. This may have led to the development of a communication style characterised by shorter sentences and less convoluted syntactic constructions. These changes may have been initiated by second language learners themselves (Evenks and Russians learning the language) through simplifying and overgeneralising during the language learning process, but they also may have been stimulated or perhaps reinforced by native Sakha/Dolgan speakers in an attempt to accommodate to the linguistic abilities of non-native speakers (i.e. foreigner talk).

9.4.2 INTERPRETATION OF CHANGES USING LANGUAGE CONTACT THEORY, SOCIO-HISTORICAL AND GENETIC INFORMATION

In addition to lexical copies from Russian, there are significant structural differences between Dolgan and Sakha that developed as a result of contact with Russians. As for the contact setting with the Evenks, the main process underlying these structural changes was identified as imposition due to language shift. However, an important difference between the two situations concerns the direction of the shift and the agents of change. In the contact setting with the Evenks it was the Evenks who shifted towards Dolgan and imposed structures from their dominant L1 (Evenki) onto their L2 (Dolgan). In contrast, in the Russian contact setting it is the Dolgans *themselves* who cause change to their traditional language. The on-going shift to Russian in the Dolgan community induces the Dolgans to impose structures from their new dominant language (Russian) onto their non-dominant language (Dolgan). Since Russian dominance can only be safely assumed in the youngest one or two generations, the structural changes described in this thesis must have come about only recently. The fact that SVO word order and Russian-inspired clause combining structures are used by people of all ages does not contradict this hypothesis. It was shown that at the level of the individual speaker the same linguistic result can be explained by different

linguistic processes. Taking into account the internal heterogeneity of the current Dolgan speech community, a subdivision was made between Dolgan-dominant and Russian-dominant speakers, which typically correlates with age. This is the result of the language shift that is swiftly progressing through the Dolgan community, and which leads to a population that diverges with respect to their linguistic dominance, depending on the social setting (and the time) in which they grew up.

The oldest age group (> 70) can be confidently said to be dominant in Dolgan, which is also their L1, whereas Russian is their non-dominant L2. On the other hand, the youngest age group (< 40 years) is dominant in Russian, which is also typically their L1 (with the exception of children in Syndassko and Sopochnoe), whereas Dolgan is their non-dominant language. It needs to be mentioned that for the older people within this age group Dolgan may still be their L1, since they used this language with their parents during their pre-school years. However, the social environment of Russian schooling and a society that has been increasingly dominated by Russian culture and language engendered constant activation of Russian (and deactivation of Dolgan), which over time rendered Russian the dominant language, whereas Dolgan has become a distant second. The age group in-between shows too much internal variation to be classified in either category.

While the structural changes were probably initiated by members of the youngest age group through the process of imposition, it was argued that the older, Dolgan-dominant generation came to the same linguistic results through the process of borrowing. After all, both age groups found themselves in a situation of intense cultural and linguistic contact with Russians, and the constant exposure of the older generation to the Russian language, as well as to the language variety of the younger generation, may well have had its impact on the grammar of this age group too.

The idea that it is the Dolgans themselves who are fostering this variation and not Russians who learn Dolgan as a second language is clearly supported by socio-historical as well as genetic results. First, we have no historical evidence that during the last century there was a significant number of Russians who learned the Dolgan language, so therefore linguistic changes imposed by Russians can be excluded. As pointed out in the previous section, the genetic profile of the Dolgans also does not present any reason to doubt this conclusion. While some genetic material associated with the Russian population is present in the Dolgan paternal genepool, its proportion is too small to make the idea of a language shift an appealing one.

A parallel development underlying the on-going changes in Dolgan is language attrition. The constant activation of Russian and deactivation of Dolgan in the bilingual population reduces the linguistic variety of actively used forms and constructions and causes certain details of Dolgan to fall into disuse. The generally meagre exposure to Dolgan of the children today does not regenerate the latent language properties in this group and is leading to the loosening of certain grammatical rules and thus to even more loss of linguistic detail in the next generation. An example of this was the less frequent use of morphosyntactically complex relative clauses in Dolgan and the inter- and sometimes even intrapersonal morphological variation in these constructions when they were elicited. However, this applies equally to the lexical domain.

Despite Thomason's advice to concentrate on the presence rather than the absence of contact-induced changes, the question arises why these structural changes are present in Dolgan, but not in Sakha. Russian influence extends all across Siberia and is by no means limited to the Taimyr Peninsula, but clearly not all communities respond in the same way. While explanations must remain speculative, a brief exposition of some social and historical differences between the relation of the Russians with the Sakha on the one hand and with the Dolgans on the other, may help to understand the course of developments.

One very practical reason could be that the Russians simply did not reach every corner of the vast territory inhabited by the Sakha, especially the remote rural areas (Pakendorf p.c.). While in the initial stages of Russian contact the Sakha inhabited a much smaller area in central Yakutia, between the rivers Lena, Aldan and Amga (Dolgikh 1960: 377), by the time the Soviets started their large-scale transformations they had spread out over a vast area of more than 3,000,000 km² (Safronov 2000: 11). The remoteness and inaccessibility of the Siberian taiga and tundra may have made it hard for the Russian government to effectively transform society in these regions (Ivanov: 370). This is of course not to imply that this job was easy on the Taimyr. On the contrary, some places may be even harder to reach due to the harsh climate. Nonetheless the Taimyr, which is today an area of 879,000 km², is considerably smaller than the Sakha Republic, and the populated areas were mainly concentrated along the rivers.

However, more important factors may include the attitude of the population towards the Russians and contrasts in lifestyle and mode of subsistence before and after their arrival. While these factors are mainly socio-historical, they influence the relation between the indigenous population and the Russians and may well

have had repercussions on language, and language attitude. As far as lifestyle is concerned, the Sakha traditionally had a semi-settled pastoralist mode of subsistence, with predominantly stationary dwellings, one for summer and one for winter. While the arrival of the Russians and the creation of planned villages (*poselkovaniye*) forced them to live in a different kind of house, become more fixed to one place and to live in a more densely populated setting than they were used to, they could continue their traditional occupation of pastoralism and hay-making in much the same way as they had done before. The Dolgans, on the other hand, used to lead a predominantly nomadic life as reindeer herders and traders, without permanent settlement. While a camp would often consist of more than one family, especially in the summer, the settlements would always be temporary for two weeks at the most, after which they would move to new pastures for the reindeer herd. For the Dolgans, the transition to permanent settlement would not only have meant a change in house type and population density, but also a major change in lifestyle (nomadic to settled) and mode of subsistence and occupation (reindeer herding to working on a *kolkhoz*). While reindeer brigades continued to exist, most people were involved in work on the *kolkhoz* and in the village. Importantly, these changes made the Dolgans highly dependent on the Russians. The concept of village life and the new professions were unknown to them so interaction with the Russians and proficiency in the Russian language was crucial to acquire the new ways. Education in Russian and a prohibition to speak their native tongue of course enhanced this trend. The increased dependency on the Russians in combination with a strong everyday confirmation that knowledge of Russian was essential to get on in the new lifestyle, may have contributed to a general attitude associating the Russian language with progress and usefulness, whereas Dolgan became more and more associated with communication in small circles and traditional settings.

Finally, contact with the Russians has played an important role in the formation of the Dolgans as a separate ethnolinguistic group. As traders along the Khatanga Trading Way, interaction with the Russians seems to have been an integrated aspect of Dolgan life, which may have made their community more open towards the increasing Russian influence than the Sakha communities.

9.4.3 SUMMARY

The analysis of Russian-induced differences between Dolgan and Sakha has demonstrated that the observed morphosyntactic variation began to develop parallel to the emergence of Russian-Dolgan bilingualism in the Dolgan speech community. This trend intensified most strikingly during the second half of the 20th century and is still continuing, foreshadowing a complete shift to Russian within the next few generations if no active measures are taken. Considering the fact that the growing dominance of Russian in the Dolgan speech community is paralleled by an increasing deactivation of Dolgan, imposition of Russian features onto Dolgan is accompanied by signs of language attrition.

9.5 THE USE OF DOLGAN/SAKHA AS A LINGUA FRANCA

Throughout the above discussion, the use of Dolgan/Sakha as a lingua franca has been mentioned repeatedly as an additional explanation for the nature and the geographical distribution of certain differences between modern Dolgan and Sakha. It was associated in particular with regularisation and with the observed preference for paratactic structures over relative clauses to express complex propositions, as well as with the morphological simplification of relative clauses, if they occur. It is important to point out from the beginning that there is considerable overlap between a scenario of shift and a situation where the recipient language is used as a lingua franca, and that probably no clear distinction can be made as far as the cognitive processes of the individual L2 learner are concerned. In both situations the novel use of the recipient language is initiated by L2 speakers, and therefore it is primarily driven by principles of second language learning (see Section 3.1.3). Nevertheless, there are a number of fundamental differences, mostly of a sociolinguistic nature, which is why the use as a lingua franca deserves its dedicated space.

First, in contrast to a shift situation it may not be possible to identify a single source language for the linguistic differences that set a lingua franca apart from its L1 variety. One characteristic of a lingua franca is its use between peoples with various linguistic backgrounds, and they may all bring in features from their L1. Thus, while the lingua franca itself is unambiguously the recipient language, the diverse origins of lingua franca speakers make it impossible to single out one

source language for the changes. In the case of Dolgan, the L2 speakers were predominantly Evenks, but also Russians and Samoyedic people (Nenets, Enets and Nganasan) participated in the trade along the Khatanga Trading Way, albeit to a lesser extent. Since the aim of a lingua franca is to facilitate communication among speakers from different linguistic backgrounds, one can imagine that structures which are very specific to a speaker's L1 will not so easily become established as a feature of the lingua franca. If speaker A with dominant language L1A imposes a feature on the lingua franca that is uncommon in the area and cross-linguistically marked, it may not be easily interpreted by speaker B with dominant language L1B, in which this feature is not present, and may therefore not take root in the lingua franca. On the other hand, imposed features that are common in the area, or that are commonly observed in L2 learning (such as regularisation) are more likely to become accepted by the lingua franca-using community. Therefore, features in a lingua franca that do not match a specific source language, and that are common in second language acquisition are plausible candidates for an explanation in terms of intergroup communication.

To apply this to Dolgan, the abovementioned regularised forms and morphosyntactic simplification do not match any language specific structures of Evenki, Russian or Samoyedic languages, and can not be traced back to a particular source language. Rather these changes developed as a result of general language learning principles, and may therefore be attributed to the function of Dolgan/Sakha as a lingua franca. This does of course not exclude the possibility that Evenki speakers played an important part in the rise of these changes, especially when bearing in mind that the majority of L2 learners probably consisted of Evenks, but these data in isolation do not provide direct evidence and so this development should not be limited to this group.

Second, a lingua franca is not only used as a means of communication between L1 speakers and L2 speakers, as is the case in a typical contact situation, but also among L2 speakers themselves. In trading situations, Dolgan/Sakha was used between Sakha people and other groups, but also among these other groups themselves, none of which were L1 speakers of Dolgan/Sakha³. This suggests that in this setting, input from L1 speakers may have been sparse, or even absent, at times, leaving more room for the development and establishment of innovative use by L2 speakers.

³ Of course Taimyr Pidgin also fulfilled this function, but its use seems to have been much more restricted, and was limited mainly to communication between Dolgans and Nganasans.

Finally, there is reason to assume that in intergroup communication not only L2 speakers but also L1 speakers make modifications to the standard variety of their language, thus contributing to a new language variety themselves. Since an easily understandable message is beneficial for everyone, a socially empathic L1 speaker may use the most transparent and iconic ways to get the message across (see foreigner talk, mentioned in Section 9.4.1). While this fact does not account for regularisation (even for the most empathic native speakers it would go too far to eliminate irregularities in inflectional paradigms to accommodate to L2 speakers), it does explain the predominance of shorter and morphosyntactically less complex structures in Dolgan when compared to Sakha.

In this context it needs to be mentioned that the Taimyr was not the only region where Sakha was used as a lingua franca. According to Wurm (1996: 971), this phenomenon was rather widespread in other parts of Siberia as well. Therefore one may ask why on the Taimyr, and not so much in other areas, the contact variety became accepted among L2 as well as L1 speakers, resulting in the language called Dolgan today. While explanations must remain speculative, an important difference between the situation on the Taimyr and other contact settings of Sakha is the substantial amount of genetic admixture between L1 speakers (Sakha) and L2 speakers (mostly Evenks) of Sakha. If in most other contact situations the use of Sakha was restricted to the domain of trade and intergroup communication, the common interethnic marriages on the Taimyr, reflected by this admixture, brought the lingua franca variety to peoples homes as well. This transfer of the lingua franca into the domestic sphere probably highlighted the need for a common language even more and may have facilitated language shift. Since Dolgan/Sakha was the lingua franca anyway, the most likely direction of shift in these mixed families was for Evenks to shift to Dolgan/Sakha rather than the other way around.

Thus the combination of Dolgan/Sakha as a lingua franca (associated with generalisations and morphosyntactic simplification), with the fact that most of the shifting people were Evenks (associated with generalisations and with changes directly mirroring Evenki structures), provides a plausible framework in which to view the range of observed differences between Dolgan and Sakha.

9.6 CONCLUSION

By recognising the linguistic and demographic variation within the community under consideration, this study confirms the importance of multicausality for explanations of contact-induced change. It shows how one linguistic outcome can develop as a result of multiple linguistic processes depending on group-internal differences in linguistic dominance within the bilingual population as well as on the function of the language in the sociolinguistic landscape. It also shows that, contrary to what some scholars propose (cf. Lucas 2012), a person's L1 need not necessarily coincide with his dominant language. Although many Dolgans who grew up in the 1950's and 1960's learned Dolgan as their L1 from their parents, many of them are now dominant in Russian, and admit that they are forgetting their native tongue.

In contrast to the contact situation between Dolgans and Evenks, which belongs to the past, the contact setting with Russians is still on-going and thus provides a real-time study of a contact situation for which social details are still available. Therefore it is an important contribution to the collection of case studies that is needed to gain insights into correlations between social settings and their linguistic outcomes, and can thus contribute to the further development of language contact theory, as well as test existing theories.

CHAPTER 10 CONCLUDING REMARKS AND FUTURE OUTLOOK

The research for this thesis was conducted with two primary aims: a) to investigate contact-induced change in Dolgan; and b) to get a better insight into the origins and history of the Dolgan. The first goal was achieved by investigating a selection of linguistic consequences of two different contact situations that have left their vestiges in the current Dolgan language, namely that with Evenki and that with Russian. It was concluded that in both contact settings the observed changes are motivated primarily by the linguistic process of imposition as a result of language shift. However, an important difference between them concerns the direction of the shift and the agents of change. In the contact setting with Evenki there was a shift *towards* the Dolgan language. The Tungusic Evenks, who did not speak a Turkic language, shifted to Dolgan/Sakha and imposed structures from their L1 (Evenki) onto their L2 (Dolgan/Sakha) due to imperfect learning. In contrast, the shift referred to in the contact setting between Russians and Dolgans is a shift *away from* Dolgan, and is initiated by the Dolgan people themselves. The structural changes are the result of a change in linguistic dominance of the Dolgans, who now impose linguistic properties from their dominant language Russian onto Dolgan. While technically the underlying process, as well as the character of the linguistic consequences, may be comparable, there is an obvious social difference: in the contact setting with the Evenks, the shift and the

consequent imposed changes are the result of the increased social dominance of Dolgan, whereas in the second case they instead reflect a decrease.

During the climax of contact between Evenks and Dolgans, which took place between the 18th and 19th centuries along the Khatanga Trading Way, the Dolgan people occupied a socially dominant position in this niche. Over time, their language acquired the function of a *lingua franca* and served as an important instrument of intergroup communication across all ethnic groups that participated in the trade. This in turn produced more and more second language speakers, and may have been an important motivation for other Evenk clans, who were the other main figures in the Taimyr trade, to shift completely to the Dolgan language.

In contrast, the contact situation with the Russians is marked by a decrease in the social significance of Dolgan. The ideological invasion of the Russians during the 20th century, which had serious social, political and cultural repercussions, was paralleled by a linguistic invasion as well. The initial consequence was widespread and relatively balanced bilingualism within the Dolgan speech community, but toward the 1970's the balance of linguistic dominance tipped towards the Russian side. By now the majority of Dolgans is dominant in Russian, and the linguistic consequences of this dominance are felt as structural changes in Dolgan, based on the model of Russian. Thus, within less than a century many Dolgans have become non-dominant speakers, or even L2 speakers, of their traditional language, and now a shift to Russian seems almost irreversible.

However, evidence from other disciplines has led to the insight that in both contact settings an explanation of the linguistic changes in terms of imposition through shift alone does not tell the full story. The 'mismatch' in the Dolgan community between their Tungusic name and their Turkic language stimulates the idea of a shift, either of language or of ethnic identity. The contact-induced structural changes discussed in this thesis, in combination with the high proportion of Tungusic-associated haplogroups in the Dolgan Y-chromosome, provide strong support for the first hypothesis. However, zooming out from the individual changes, this account is not entirely satisfactory. If the Dolgans were exclusively Tungusic clans that adopted the Dolgan/Sakha language, the total amount of traces from their mother tongue is surprisingly limited. While the observed contact-induced changes are informative on their own account, within the overall picture of the Dolgan/Sakha language they constitute only a small part, suggesting that the history of the ancestors of the Dolgans has more than one

storyline. Once again it is the genetic information that proves a useful guide. Besides the impressive proportion of Tungusic-related haplogroups, nearly 40% of the paternal genepool is shared with the Turkic Sakha. The most plausible conclusion to be drawn from this is that the ethnolinguistic group identifying as Dolgans today developed in part from the fusion of Tungusic Evenk clans who shifted to Dolgan/Sakha, and in part from a significant number of Sakha people, who kept their language but shifted ethnic identity. Thus, the continuous presence of L1 Sakha-speaking people in the emerging Dolgan community would even out some of the Evenki influence that went hand in hand with the large number of L2 Sakha speakers. Therefore, recognition of these two sides of the coin, based on a combination of linguistic, genetic and historical information, accounts for the characteristic, but nonetheless relatively limited amount of Tungusic influence in the Dolgan language as it is spoken today.

A comparable conclusion can be drawn regarding the contact between Dolgans and Russians. Imposition through shift is, in this case, the main explanation for the structural variation in Dolgan. However, a careful consideration of recent history, and the social setting in which the current shift is taking place, shows that on an individual level a distinction needs to be made between Dolgan-dominant and Russian-dominant speakers in order to give a complete account of the linguistic processes that underlie the changes. While this division is anything but definite, for the sake of convenience dominance was correlated with age, where the older generation is typically dominant in Dolgan, and the younger generation dominant in Russian. The consequence of this recognition is that the same linguistic outcome (structural change) is explained through multiple motivations; by imposition through shift in the Russian-dominant part of the population, and simultaneously by borrowing in an intense contact situation in the Dolgan-dominant part of the population.

Finally, the genetic information on the Dolgan population was used to conclusively exclude the possibility that these structural changes occurred through imposition by Russians who shifted to Dolgan. In accordance with the historical information on this issue, the limited amount of Russian-related haplogroups in the Dolgan population clearly confirm this idea.

As stated in the introduction, one of the incentives for reopening the investigation into the origins of the Dolgans was the contradictory information in the ethnographic, historical and linguistic sources, and the resulting confusion about the way in which their 'mixed' heritage should be understood. The research

presented here has shown that careful consideration of data from all these sources provides a more detailed and realistic image of the undocumented past. More importantly, it illustrates that a multidisciplinary approach and a synthesis of the outcomes from different research areas can yield new insights, and add meaning to data that seem insignificant or contradictory when considered individually.

While I am convinced that this is true in almost any research discipline, it is particularly apparent in the study of language and even more so of contact-induced change, which is tightly interwoven with other aspects of the human individual, as well as with human society, and thus touches upon fields of study including psychology, cognitive sciences, anthropology and sociology. Since such fields rely heavily on interpretation of ephemeral aspects of human life, genetic information provides a more permanent and objective frame for the interpretation of the linguistic material. Therefore, the combination of insights from the abovementioned fields, and their repercussions on language use, greatly advances our understanding of the contact situation as a whole.

This thesis also provides supporting evidence for the idea that contact-induced change often needs to be explained by multiple motivations. If one wants to do justice to the complexity of a contact situation, it is often not possible to hold on to a binary dichotomy between borrowing or imposition, and it is necessary to recognise that a single linguistic change may be motivated by more than one process of change.

Finally, it needs to be emphasised that this thesis has not only given answers, but it has also generated questions regarding the history and language of the Dolgans, contact-induced change, as well as the further development of theories of language contact. Linguistically, it has exposed the lack of phonetic and intonational data on Dolgan, as well as on its neighbouring languages. Since the judgement of native Dolgan speakers, and even of my non-native ear, suggest that there are differences in the phonetics, and particularly in the intonation patterns of Dolgan and Sakha, this domain is begging for systematic research. This would involve investigation of intonation patterns in different sentence types (questions, declarative sentences, etc.) in spontaneous speech, as well as in elicitation. Spontaneous texts are essential to find the most natural intonation patterns, whereas elicited sentences must be included for cross-linguistic comparison to make sure that the exact same meaning is expressed. This comparison would have to take place primarily between Dolgan and Sakha (and perhaps other Turkic languages), followed by the same procedure for Evenki (and other Tungusic

languages) to scrutinise whether differences between Dolgan and Sakha are contact-induced. Besides conducting this research out of curiosity, its results would have broader implications for the conclusions about Dolgan prehistory. After all, changes in the sound system, and supposedly also intonation, are, like changes in morphosyntax, associated with imposition and possibly language shift. While the latter were shown to have taken place in Dolgan, more evidence of structural change affecting the sound system would certainly strengthen the case for reconstructing such a sociolinguistic context.

This naturally highlights a second area for future work, namely the composition of more annotated and comparable text corpora. The present study has confirmed their importance for the field of contact linguistics, as well as highlighted the amount of work that still needs to be done in this area. While grammars are a very rich source of language data, they normally do not provide information on frequency of use or on subtle semantic differences between grammatical constructions, both of which are important topics to investigate in language contact. While I am aware that the body of linguistic corpora is rapidly expanding, I feel the need to mention it nonetheless, since it is clear that an available annotated corpus of Evenki texts would give the conclusions of this research, in particular those regarding frequency of use, a stronger foundation.

While multi-causality of contact-induced change has been illustrated in many case studies, and many scholars have pointed out the significance of it, this recognition has not yet been sufficient to formulate a theory of language contact that is able to incorporate all these factors in a systematic way. Admittedly, the complexity of a language contact situation makes this a very daunting task indeed. The large number of variables, and the countless ways in which they may interact, will probably never allow us to develop a theory that is able to predict flawlessly linguistic outcomes from social contexts. However, while this aspiration may verge on the impossible, there is no reason why we should not try to approach this goal as closely as we can. One important initiative for the advancement of language contact theory would be the reorganisation and improved accessibility of the available material on the subject. Within the last fifty years an abundance of individual case studies has emerged for many contact situations across the world. While some are more detailed than others, they all provide data that is of crucial value for the study of contact-induced change in individual contact situations, as well as for the development of models and theories of language contact. However, in their current shape we are not able to exhaust their value to the full. They are

scattered across the literature, and since there has never been one comprehensive theoretical framework for contact-induced change, they often use different criteria and different terminology to describe the changes and their underlying motivations, which makes them hard to compare.

One way to improve this situation would be to collect all the individual case studies of contact-induced change and compile them as a searchable database. If they are analysed and coded for a set of linguistic, sociological, historical, psychological and ethnographical criteria in a cross-linguistically valid way, it would be possible to uncover correlations between social factors and their linguistic outcomes cross-linguistically in a quantitative, as well as qualitative way. While on the one hand such a database would provide a rich and detailed data source and a great body of reference for contact linguists working on individual case studies, on the other hand it would enable us to abstract away from individual cases, thus coming one more step closer to a comprehensive theory of language contact.

It is obvious that the picture of the Dolgan people and their history as represented in this thesis is far from complete. Nonetheless, I hope to have contributed to a clearer understanding of their history and to have eliminated some of the more extreme views on their identity, as well as on their language. In addition, by combining socio-historical, linguistic and genetic data, I hope to have added shape, detail and colour to the black and white sketch that Dolgan history was so far.

REFERENCES

- Afanasyev, P.S., Voronkin, M.S., Alekseev, M.P. 1976. *Dialektologicheskii slovar' yakutskogo yazyka: sodержit svyshe 85000 slov* [Dialectal dictionary of the Yakut language: contains over 85,000 words]. Moskva: Nauka.
- Aikhenvald, A.Y. 2003. Mechanisms of change in areal diffusion: new morphology and language contact. *Journal of Linguistics* 39 (1):1-29.
- Aikhenvald, A.Y. 2006. Reflections on language contact, areal diffusion, and mechanisms of linguistic change. In *Sprachbund in the West African Sahel. Afrique et Langage*, edited by B. Caron and Zima, P. Paris: Peeters.
- Aitchison, J. 1981. *Language change: progress or decay?* London: Fontana.
- Aksenova, E.E., Beltyukova, N.P., Kosheverova, T.M. 1992. *Solvar' dolgansko-russkiy i russko-dolganskiy* [Dictionary Dolgan-Russian and Russian-Dolgan]. St. Petersburg: Prosvesh'enie.
- Altenberg, E. 1991. Assessing first language vulnerability to attrition. In *First Language Attrition*, edited by R. Seliger, and Vago, R. Cambridge: Cambridge University Press.
- Andersen, H. 1988. Centre and periphery: adoption, diffusion and spread. In *Historical dialectology*, edited by J. Frisiak. Berlin: Mouton de Gruyter.
- Anderson, D. G. 2000. *Identity and ecology in arctic Siberia*. New York: Oxford University Press.
- Anderson, G.D.S. 1998. *Xakas*. Vol. 251. München: Lincom Europa.
- Anderson, G. D. S. 2006. Towards a typology of the Siberian linguistic area. In *Linguistic areas: convergence in historical and typological perspective*, edited by Y. Matras, and McMahon, A., and Vincent, N. Basingstoke: Palgrave Macmillan.
- Anikin, A. E. 2003. *Etimologicheskii slovar' russkikh zaimstvovaniy v yazykakh Sibiri* [Etymological dictionary of Russian loan words in the languages of Siberia]. Novosibirsk: Nauka.
- Aralova, N.B. 2007. *Dannye s mordovskoy ekspeditsii* [Data from an expedition to Mordovia]. Unpublished fielddata.
- Arkheograficheskaya komissiya. 1875. *Russkaya istoricheskaya biblioteka*. Vol. 2. St. Petersburg.

- Arkheograficheskaya komissiya. *Dopolneniya k Aktam istoricheskim*. Vol. 2. St. Petersburg.
- Arkhipov, A.V. 2009. *Tipologiya komitativnykh konstrukciy [A typology of comitative constructions]*. Moscow: Znak.
- Artemyev, N.M. 2001a. *Dolganskiy yazyk, chast' 1 [The Dolgan language, part 1]*. Saint Petersburg: RGPU im. A.I. Gerzena.
- . 2001b. *Dolganskiy yazyk, chast' 2 [The Dolgan language, part 2]*. Saint Petersburg: RGPU im. A.I. Gerzena.
- Avrorin, V.A. 1961. *Grammatika nanayskogo yazyka [A grammar of Nanai]*. Moscow: Izdatel'stvo Akademii Nauk SSSR.
- Behaghel, O. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25:110-142.
- Benzing, J. 1955. *Lamutische Grammatik*. Wiesbaden: Steiner.
- . 1956. *Die tungusischen Sprachen, Versuch einer vergleichenden Grammatik*. Wiesbaden: Franz Steiner Verlag.
- Berta, A. 1998. Middle Kipchak. In *The Turkic languages*, edited by L. Johanson and Csató, E. London - New York: Routledge.
- Bettu, E.S. 2010. *Imena dolgan [Dolgan names]*. Krasnoyarsk: Izdatel'skiy dom 'Novyy Yenisey'.
- Betz, W. 1949. *Deutsch und Lateinisch; die Lehnbildungen der althochdeutschen Benediktinerregel*. Bonn: H. Bouvier.
- Bialystok, E., Craik, F.I.M., Luk, G. 2012. Bilingualism: consequences for mind and brain. *Trends in cognitive sciences* 16 (4):240-250.
- Blank, A. 1997. Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen. *Zeitschrift für romanische Philologie* XVI (285).
- . 2003. Words and concepts in time: towards diachronic cognitive onomasiology. In *Words in Time: diachronic semantics*, edited by R. Eckhardt, Van Heusinger, K. and Schwarze, C. Berlin - New York.
- Böhtlingk, O. N. [1851] 1997. *Über die Sprache der Yakuten*. Vol. 35, *Uralic and Altaic Series*. The Hague: Mouton & Co.
- Boldyrev, B.V. 1994. *Russko-evenkiyskiy slovar' [Russian-Evenki dictionary]*. Novosibirsk: Nauka.
- . 2007. *Morfologiya evenkiyskogo yazyka [Morphology of the Evenki language]*. Novosibirsk: Nauka.
- Bolinger, D. 1975. *Aspects of language*. second ed. New York: Harcourt Brace Jovanovich.

- Brown, C.H. and Witkowski, S.R. 1983. Polysemy, lexical change and cultural importance. *Man* 18 (1):72-89.
- Bulatova, N. and Grenoble, L. 1999. *Evenki*. München ; Newcastle: LINCOM Europa.
- Butters, R.R. 2001. Chance as a cause of language variation and change. *Journal of English linguistics* (29): 201-213.
- Bybee, J. 1991. Natural morphology: the organization of paradigms and language acquisition. In *Cross currents in second language acquisition and linguistic theories*, edited by T. Huebner and Ferguson, C.A. Amsterdam - Philadelphia: Benjamins.
- Bybee, J. and Perkins, R. and Pagliuca, W. 1994. *The evolution of grammar: tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.
- Campbell, L. 1987. Syntactic change in Pilpil. *International journal of American linguistics* 53 (3):253-280.
- Castrén, M.A. 1856. *Reiseberichte und Briefe aus den Jahren 1845-1849*, edited by A. Schiefner.
- . 1860. Puteshestvie v Sibir' 1845-1849 [Journey to Siberia]. *Magazin zemlevedeniya i puteshestvii* 6 (2):199-482.
- Chafe, W. 1980. *The Pear Stories: cognitive, cultural, and linguistic aspects of narrative production*. Vol. 3, *Advances in discourse processes*. Norwood N.J.: Ablex Pub.
- Cheremisina, M. I. 1995. *Grammatika sovremennogo yakutskogo yazyka: syntaksis [A grammar of the modern Yakut language: syntax]*. Novosibirsk: Nauka, Sibirskaya izdatel'skaya firma RAN.
- Cincius, V.I. 1982. *Negidal'skiy yazyk: issledovaniya i materialy [the Negidal language: investigations and materials]*. Leningrad: Nauka.
- Clyne, M. 1992. Linguistic and sociolinguistic aspects of language contact, maintenance and loss: towards a multifacet theory. In *Maintenance and loss of minority languages*, edited by W. Fase, Jaspaert, K. and Kroon, S. Amsterdam: John Benjamins.
- Comrie, B. 1976. *Aspect, an introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
- . 1989. *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.
- . 2008. Subordination, coordination: form, semantics, pragmatics - setting the scene. In *Subordination and coordination strategies in North Asian languages*, edited by E. Vajda. Philadelphia: Benjamins.

- . [1985] 2000. *Tense*. Cambridge: Cambridge University Press.
- Cristofaro, S. 2003. *Subordination, Oxford Studies in Typology and Linguistic Theory*. Oxford: Oxford University Press.
- Croft, W. 2003. *Typology and universals*. second edition, *Cambridge textbooks in linguistics*. Cambridge - New York: Cambridge University Press.
- . 2003a. Mixed languages and acts of identity: an evolutionary approach. In *The mixed language debate*, edited by Y. Matras and Bakker, P. Berlin: Mouton de Gruyter.
- Crowley, T. 2000. Simplicity, complexity, emblematicity and grammatical change. In *Processes of language contact, studies from Australia and the South Pacific*, edited by J. Siegel. Québec: Fides.
- Csató, E. and Johanson, L. 1998. Turkish. In *The Turkic languages*, edited by L. Johanson and Csató, E. London - New York: Routledge.
- Dahl, Ö. 1985. *Tense and aspect systems*. New York: Blackwell.
- Dawkins, R. M. and Halliday, W.R. 1916. *Modern Greek in Asia Minor: a study of the dialects of Sílli, Cappadocia and Phárasa, with grammar, texts, translations and glossary*. Cambridge.
- De Jong, F. 2007. *A grammar of written Uyghur*. Utrecht: Houtsma.
- Denisov, V. 2008. *Khronologiya Taymyra; khronologiya istoricheskikh sobytiy genezisa Taymyra i noril'skogo rayona [Chronology of the Taimyr: the chronology of historical events concerning the development of the Taimyr and the Norilsk region]*. Noril'sk: APEX.
- Diessel, H. and Tomasello, M. 2001. The acquisition of finite complement clauses in English: a usage based approach to the development of grammatical constructions. *Cognitive Linguistics* 12:97-141.
- Dik, S. C. 1997. *The theory of functional grammar: the structure of the clause*. Berlin: Mouton de Gruyter.
- Doerfer, G. 2004. Etymologisches-Ethnologisches Wörterbuch tungusischer Dialekte (vornehmlich der Mandschurei). Hildesheim, Zürich - New York: Georg Olms Verlag.
- Dolgikh, B.O. 1929. Naseleniye poluostrova Taymyr i privileyush'ego k nemu rayona [The population of the Taimyr Peninsula and the adjacent region]. *Severnaya Asia* 2:49-76.
- . 1960. *Rodovoy i plemennoy sostav narodov sibiri v XVII veke [Kin and clan structure of the peoples of Siberia in the 17th century]*. Moscow: Izdatel'stvo Akademii Nauk SSSR.

- . 1963. Proiskhozhdeniye dolgan [the origin of the Dolgans]. In *Sibirskiy etnograficheskiy sbornik*. Moscow: Izdatel'stvo Akademii Nauk SSSR.
- Erdal, M. 1991. *Old Turkic word formation: a functional approach to the lexicon*. Vol. 7-8. Wiesbaden: Harrassowitz.
- . 1998. Old Turkic. In *The Turkic languages*, edited by L. Johanson and Csató, E. London - New York: Routledge.
- . 2004. *A grammar of Old Turkic*. Leiden - Boston: Brill.
- Fabricius-Hansen, C. and Ramm, W. 2008. 'Subordination' vs. 'Coordination' in sentence and text: a cross-linguistic perspective. Philadelphia: Benjamins.
- Forston, B. W. 2003. An approach to semantic change. In *The Handbook of Historical Linguistics*, edited by B. D. Joseph and Janda. R.D. Oxford: Blackwell Publishing.
- Forsyth, J. 1992. *A history of the peoples of Siberia: Russia's north Asian colony 1581-1990*. Cambridge, UK: Cambridge University Press.
- Frazier, L. and Rayner, K. 1988. Parametrizing the language processing system: left- vs. right-branching within and across languages. In *Explaining language universals*, edited by J. A. Hawkins. Oxford: Blackwell.
- Friedman, V.A. 2003. *Turkish in Macedonia and beyond: studies in contact, typology, and other phenomena in the Balkans and the Caucasus*. Wiesbaden: Harrassowitz.
- Geeraerts, D. 2010. *Theories of lexical semantics*. Oxford ; New York: Oxford University Press.
- Givón, T. 1990. *Syntax: a functional-typological introduction*. Amsterdam and Philadelphia: John Benjamins.
- Grosjean, F. 2000. The bilingual's language modes. In *The bilingualism reader*, edited by L. Wei. London - New York: Routledge.
- . [2000] 2007. The bilingual's language modes. In *The bilingualism reader*, edited by L. Wei. London and New York: Routledge.
- Gumperz, J. and Wilson, R. 1971. Convergence and creolization: a case from the Indo-Aryan/Dravidian border. In *Pidginization and creolization of languages*, edited by D. Hymes. Cambridge: Cambridge University Press.
- Gurvich, I.S. 1966. *Etnicheskaya istoriya Severo-Vostoka Sibiri [The ethnic history of the North-East of Siberia]*. Vol. 89, AN SSSR, *Trudy instituta etnografii im. N.N. Miklukho-Maklaia. Novaya seria*. Moscow: Akademia Nauk SSSR.
- Haiman, J. and Thompson, S.A., ed. 1988. *Clause combining in grammar and discourse*. Amsterdam: Benjamins.

- Hamers, J.F. and Blanc, M.H.A., ed. 2000. *Bilinguality and Bilingualism*. Cambridge: Cambridge University Press.
- Harris, A.C. and Campbell, L. 1995. *Historical syntax in cross-linguistic perspective, Cambridge studies in linguistics*. Cambridge ; New York: Cambridge University Press.
- Haspelmath, M. 2004. *Coordinating constructions*. Amsterdam: Benjamins.
- Haspelmath, Martin and Tadmor, U. 2009. *Loanwords in the world's languages : a comparative handbook*. Berlin, Germany: De Gruyter Mouton.
- Haugen, E. I. 1949. Problems of bilingualism. *Lingua* 2:271-290.
- . 1950. The analysis of linguistic borrowing. *Language* 16:210-231.
- . 1953. *The Norwegian language in America; a study in bilingual behavior*. 2 vols, *Publications of the American Institute, University of Oslo in coöperation with the Department of American Civilization, Graduate School of Arts and Sciences, University of Pennsylvania*. Philadelphia: University of Pennsylvania Press.
- Hawking, S. 1988. *A brief history of time*. New York: Bantam Books.
- Hawkins, J.A. 1983. *Word order universals*. San Diego: Academic Press.
- Heath, J. 1978. Linguistic diffusion in Arnhem Land. *Australian Aboriginal Studies Research and Regional Studies* (13).
- Heine, B. 2008. Contact-induced word order change without word order change. In *Language contact and contact languages*, edited by P. Siemund, Kintana, N. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Heine, B, and Kuteva, T. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Hickey, R. 2010. *The handbook of language contact, Blackwell handbooks in linguistics*. Chichester, West Sussex - Malden, MA: Wiley-Blackwell.
- Hock, H. H. 1986. *Principles of historical linguistics*. Berlin: Mouton de Gruyter.
- Hock, H. H and Joseph, B.D. 1996. *Language history, language change and language relationship: an introduction to historical and comparative linguistics*. Berlin: De Gruyter.
- Ivanov, V. N. 2007. *Yakutiya, istoriko-kul'turnyy atlas [Yakutia: historical-cultural atlas]*. Moskva: Feorija.
- Jakobson, R. [1938] 1962. Sur la théorie des affinités phonologiques entre des langues. In *Selected writings*. The Hague: Mouton.
- Janhunen, J. 2003. *The Mongolic languages*. Vol. 5. London: Routledge Curzon.

- Janse, M. 1998. Grammaticalization and typological change: the clitic cline in Inner Asia Minor Greek. In *Productivity and creativity: studies in general and descriptive linguistics in honor of E.M. Uhlenbeck*, edited by M. Janse. Berlin: De Gruyter.
- Johanson, L. ed. 1992. *Strukturelle Faktoren in Türkischen Sprachkontakten*. Vol. XXIX, *Sitzungsberichte der wissenschaftlichen Gesellschaft an der Johann Wolfgang Goethe-Universität Frankfurt am Main*. Stuttgart: Franz Steiner Verlag.
- Johanson, L. 1998. *The history of Turkic*. London and New York: Routledge.
- . 1998. The structure of Turkic. In *The Turkic languages*, edited by L. a. C. Johanson, E. London and New York: Routledge.
- . 2002a. *Structural factors in Turkic language contacts*. Translated by V. Karam. Richmond, Surrey: Curzon.
- . 2002b. Contact-induced change in a code-copying framework. In *The interplay of internal, external and extra-linguistic factors*, edited by M. C. Jones, Esch, E. Berlin - New York: Mouton de Gruyter.
- Johanson, L. and Csató, E., ed. 1998. *The Turkic languages*. London, New York: Routledge.
- Joseph, B. D. 1998. Diachronic morphology. In *The handbook of morphology*, edited by A. Z. Spencer, A.M. Oxford: Blackwell.
- Kałużynski, S. 1962. *Mongolische Elemente in der jakutischen Sprache*. Warszawa: Państwowe Wydawnictwo Naukow.
- . 1995. *Iacutica Prace jakutoznawce*. Warszawa: DIALOG.
- Karafet, T.M., Osipova, L.P., Gubina, M.A. 2002. High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Human Biology* 74 (6):761-789.
- Keenan, E.L. 1985. Relative clauses. In *Language typology and syntactic description, Vol. II: Complex constructions*, edited by T. Shopen. Cambridge: Cambridge University Press.
- Keenan, E.L. and Comrie, B. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8:63-99.
- Kerswill, P. 1996. Children, adolescents, and language change. *Language variation and change* (8):177-202.
- Kharitonov, L.N. 1947. *Sovremennyy yakutskiy yazyk. Chast' pervaya: fonetika i morfologiya* [The modern Yakut language. Part I: phonetics and phonology]. Yakutsk: Nauchno-Issledovatel'skiy Institut yazyka, literatury i istorii

YaASSr.

- . 1960. *Formy glagol'nogo vida v yakutskom yazyke* [Verbal aspect in the Yakut language]. Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- Khitrov, D.I. 1858. *Kratkaya grammatika yakutskogo yazyka* [A brief grammar of the Yakut language]. Moscow.
- Khlobystin, L.P. 1972. Absolutnyy i otnositel'nyy vozrast pamyatnika Tagenar VI [Absolute and relative age of the Tagenar VI tombstones]. In *Problemy absolutnogo datirovaniya v Arkhiologii*. Moscow.
- Khlobystin, L.P. and Gracheva G.H. 1993. *Taymyrskie kul'tury i problemy zaseleniya vostochnoy Sibiri* [Cultures of the Taimyr and the peopling of eastern Siberia]. Vol. 10, *Arkhiologicheskiye izyskaniya*.
- Kirchner, M. 1998. Kirghiz. In *The Turkic languages*, edited by E. Csátó and Johanson, L. London - New York: Routledge.
- Koch, H. 1996. Reconstruction in morphology. In *The comparative method reviewed: regularity and irregularity in language change*, edited by M. Durie and Ross, M. New York: Oxford University Press.
- Korkina, E.I. 1970. *Nakloneniya glagola v yakutskom yazyke* [Verbal inflection in the Yakut language]. Moscow: Nauka.
- Kornfilt, J. 1997. *Turkish*. London: Routledge.
- Kortmann, B. 1997. *Adverbial subordinators: A typology and history of adverbial subordinators based on European languages*. Berlin - New York: Mouton de Gruyter.
- . 1997. *Adverbial subordinators: A typology and history of adverbial subordinators based on European languages*. Berlin - New York: Mouton de Gruyter.
- Krivoshapkin, M.F. 1865. *Yeniseyskiy okrug, i ego zhizn'* [The Yenisey region and its life]. St. Petersburg: Geograficheskoye obsh'estvo.
- Krizman, J. et al. 2012. Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences* 109 (20):7877-7881.
- Labov, W. 1980. The social origins of sound change. In *Locating Language in Time and Space*, edited by W. Labov. New York: Academic Press.
- . 1994. *Principles of linguistic change: internal factors*. Oxford: Blackwell.
- Lehmann, C. 1988. Towards a typology of clause linkage. In *Clause combining in grammar and discourse*, edited by J. Haiman and Thompson, S.A. Amsterdam: Benjamins.

- Lehmann, W.P. 1992. *Historical linguistics: an introduction*. third edition. London - New York: Routledge.
- Lenneberg, E.H. 1967. *Biological foundations of language*. New York: Wiley.
- Lessing, F. D. and Haltod, M. 1995. *Mongolian-English Dictionary*. Bloomington: Mongolia Soc.
- Lightfoot, D. 1991. *How to set parameters: arguments from language change*. Cambridge: MIT Press.
- Limber, J. 1973. The genesis of complex sentences. In *Cognitive development and the acquisition of language*, edited by T. E. Moore. New York: Academic Press.
- Lucas, C. 2012. Contact-induced grammatical change: Towards an explicit account. *Diachronica* 29 (3):275-300.
- Malchukov, A. L. 1995. *Even*. München: Lincom Europa.
- Mallinson, G. and Blake, B.J. 1981. *Language typology: cross-linguistic studies in syntax*. Amsterdam: North-Holland.
- Mańczak, W. 1980. Laws of analogy. In *Historical morphology*, edited by J. Fisiak. The Hague: Mouton.
- Matic, D. and Pakendorf, B. in preparation. Non-canonical SAY in Siberia: Areal and genealogical patterns.
- Matras, Y. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36 (2):281-331.
- . 2009. *Language Contact*. Cambridge: Cambridge University Press.
- Matras, Y. and Sakel, J. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31.
- Matras, Y. and Sakel, J., ed. 2008. *Grammatical borrowing in cross-linguistic perspective: empirical approaches to language typology*. Berlin: de Gruyter.
- Matthews, S. and Yip, V. 2009. Contact-induced grammaticalization: evidence from bilingual acquisition. *Studies in Language* 33 (2):366-395.
- McConvell, P. and Meakins, F. 2005. Gurindji Kriol: a mixed language emerges from code-switching. *Australian Journal of Linguistics* (25):9-30.
- McMahon, A. M. S. 1994. *Understanding language change*. Cambridge: Cambridge University Press.
- Meillet, A. 1915. *Introduction à l'étude comparative des langues indo-européennes*. second edition. Paris: Hachette.
- . 1948. *Linguistique historique et linguistique générale*. Paris: Librairie ancienne honoré champion.

- Meuter, R.F.I. 2005. Language selection in bilinguals. In *Handbook of bilingualism, psycholinguistic approaches*, edited by J. F. Kroll and De Groot, A.M.B. Oxford: Oxford University Press.
- Middendorff, A. v. 1875. *Übersicht der Natur Nord- und Ost-Sibiriens: die Eingeborenen Sibiriens*. St. Petersburg.
- Milroy, J. and Milroy, L. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21:339-384.
- Mithun, M. 1980. A functional approach to syntactic reconstruction. In *Papers Selected from the Fourth International Conference on Historical Linguistics*, edited by E. C. Traugott. Amsterdam: John Benjamins.
- Moravcsik, E. A. 1978. Language contact. In *Universals of human language*, edited by J. H. Greenberg. Stanford: Stanford University Press.
- Muysken, P. 2010. Scenarios for language contact. In *The handbook of language contact*, edited by R. Hickey. Chichester: Blackwell.
- Myreeva, A.N. 2004. *Evenki: evenkiysko-russkiy slovar' [Evenki: Evenki-Russian dictionary]*. Novosibirsk, Russian Federation: Nauka.
- Nadkarni, M.V. 1972. NP-embedded structures in Kannada and Konkani, UCLA.
- Nedjalkov, V. P. 1997. *Evenki*. London: Routledge.
- Nettle, D. 1999. *Linguistic diversity*. Oxford: Oxford University Press.
- Newmeyer, F. J. 1998. *Language Form and Language Function*. Cambridge, Massachusetts: The MIT Press.
- Nikolaeva I. and Tolskaya, M. 2001. *A grammar of Udihe*. Vol. 22. Berlin: Mouton de Gruyter.
- O'Shanessy, C. 2005. Light Warlpiri: A new language. *Australian Journal of Linguistics* (25):31-57.
- Pakendorf, B. 2007. Contact and prehistory of the Sakha (Yakuts): linguistic and genetic perspectives, Leiden: LOT.
- . 2012. Patterns of relativization in North Asia: Towards a refined typology of prenominal participial relative clauses. In *Clause Linkage in Cross-Linguistic Perspective. Data-Driven Approaches to Cross-Clausal Syntax*, edited by V. Gast and Diessel, H. Berlin: De Gruyter Mouton.
- . Database of annotated Even texts.
- . Database of annotated Sakha texts.
- Pakendorf, B., Novgorodov, I.N., Osakovskij, V.L., Danilova, A.P., Protod'jakonov, A.P., Stoneking, M. 2006. Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Human*

Genetics 120 (3):334-353.

- Pakendorf, B. and Novgorodov, I. N. 2009. Loanwords in Sakha (Yakut), a Turkic language of Siberia. In *Loanwords in the world's languages : a comparative handbook*, edited by M. Haspelmath, Uri Tadmor. Berlin: De Gruyter.
- Pekarskij, E. K. [1907 - 1930] 1958 - 1959. *Slovar' yakutskogo yazyka [Dictionary of the Yakut language]*. Yakutsk: Yakutskiy filial Akademii Nauk SSSR.
- Penfield, W. and Roberts, L. 1959. *speech and brain-mechanisms*. Princeton: Princeton University Press.
- Popov, A.A. 1930. *Doklad orgkomitetu kompleksnoy ekspeditsii issledovaniy AN, Krasnoyarskiy Tsent Khreneniya i Izucheniya Dokumentatsii Noveyshey Istorii*.
- . 1931. *Poezdka k Dolganam [Journey to the Dolgans]*. Vol. vyp. 3-4, Sovetskaya etnografiya.
- . 2003. Dolgany: sobraniye trudov po etnografii I [The Dolgans, collection of ethnographic works I], edited by A. A. Barbolina. Saint Petersburg: Drofa.
- . 2003. Dolgany: sobraniye trudov po etnografii II [The Dolgans: collection of ethnographic works II], edited by A. A. Barbolina. Saint Petersburg: Drofa.
- Popov, N.A. and Popova, M.I. 2001. *Töröbüt tilim: uchebnik dlya 3 klassa dolganskikh shkol [My mother tongue: textbook for the 3rd class of the Dolgan schools]*. Sankt Peterburg: Prosvesh'enie.
- Poppe, N. 1926. *Uchebnaya grammatika yakutskogo yazyka [educational grammar of the Yakut language]*.
- . 1991. *Grammar of Written Mongolian*. Wiesbaden: Otto Harrassowitz.
- Radloff, W. 1908. *Die jakutische Sprache in ihrem Verhältnisse zu den Türkssprachen*.
- Raeff, M. 1956. *Siberia and the reforms of 1822*. Seattle: University of Washington Press.
- Robinson, P. and Ellis, N.C., ed. 2008. *Handbook of cognitive linguistics and second language acquisition*. New York - London: Routledge.
- Romanova, A.V. and Myreeva, A.N. 1962. *Ocherki tokkinskogo i tommotskogo govorov [Sketch of the Tokkinskij and Tommotskij dialects]*. Moskva - Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- Ross, J.R. 1973. The penthouse principle and the order of constituents. In *You take the high node and I'll take the low node*, edited by C. T. Corum, Smith-Stark T.C., Weiser, A. Chicago: Chicago Linguistic Society.
- Ross, M. 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In *The comparative method reviewed. Regularity and irregularity in language change*, edited by M. Durie and Ross, M. New York:

- Oxford University Press.
- . 2003. Diagnosing prehistoric language contact. In *Motives for language change*, edited by R. Hickey. New York: Cambridge University Press.
- . 2007. Calquing and metatypy. *Journal of Language Contact* Thema 1:116-143.
- . 2008. A history of metatypy in the Bel languages. *Journal of Language Contact* (2):149-164.
- . Forthcoming. Diagnosing contact processes from their outcomes: The importance of life stages. *Journal of Language Contact*.
- Ross, M. and Durie, M. 1996. Introduction. In *The Comparative Method Reviewed. Regularity and Irregularity in Language Change*, edited by M. Durie and Ross, M. New York: Oxford University Press.
- Rubisz, R.C., Melton, P.E. and Crawford, M.H. 2007. Molecular markers in anthropological genetic studies. In *Anthropological genetics, theory, methods and applications*, edited by M. Crawford. Cambridge: Cambridge University Press.
- Safronov, F.G., ed. 2000. *Entsiklopediya Yakutii [Encyclopedia of Yakutia]*. Moscow: Yakutskiy nauchnyj tsentr sibirskogo otdeleniya Rossiyskoy akademii nauk, Akademiya nauk Respubliki Sakha (Yakutiya).
- Sampson, G., Gil, D. and Trudgill, P., ed. 2009. *Language complexity as an evolving variable*. Oxford: Oxford University Press.
- Sandalo, F. 1995. A grammar of Kadiwéu, University of Pittsburgh, Pittsburgh.
- Sapir, E. [1921] 1970. *Language, an introduction to the study of speech*. London: Rupert Hart-Davis.
- Sauer, M. 1803. *Reise nach Sibirien, Kamtschatka und zur Untersuchung der Mündung des Kowima-Flusses, der ganzen Küste der Tschuktschen und der zwischen dem festen Lande von Asien und Amerika befindlichen Inseln. Auf Befehl der Kaiserin von Rußland, Catharina der Zweyten, in den Jahren 1785 bis 1794 unternommen vom Kapitän Joseph Billings und nach den Original-Papieren herausgegeben von Martin Sauer, Sekretär der Expedition*. Berlin - Hamburg: Aus dem Englischen übersetzt.
- Savvinov, A.I. 2005. *Dolgany: problemy etnokul'turnoy identifikatsii dolgan [The Dolgans: the ethno-cultural identification of the Dolgans]*. Novosibirsk: Nauka.
- Schmid, M. 2002. *First Language Attrition, Use and Maintenance: The case of German Jews in Anglophone countries*. Amsterdam: John Benjamins.
- Schönig, C. 1998. Turkmen. In *The Turkic languages*, edited by L. Johanson and Csató, E. London - New York: Routledge.

- Seifart, F. 2011. Bora loans in Resígaro: massive morphological and little lexical borrowing in a moribund Arawakan language. In *Cadernos de Etnolingüística*.
- Sevortyan, E.V. 1974. *Etimologicheskii slovar' tyurkskikh yazykov*. Moscow: Nauka
- Sharwood Smith, M. and Van Buren P. 1991. First language attrition and the parameter setting model. In *First language attrition*, edited by H. Seliger and Vago, R. Cambridge: Cambridge University Press.
- Sleptsov, P.A. 1972. *Yakutsko-Russkiy slovar' [Yakut-Russian dictionary]*. Moscow: izdatel'stvo sovetskaya entsiklopediya.
- Slezkine, Y. 1994. *Arctic mirrors: Russia and the small peoples of the north*. New York: Cornell University Press.
- Sloetjes H. and Wittenburg, P. 2008. Annotation by category - ELAN and ISO DCR. *Conference proceedings of the 6th international conference on language resources and evaluation (LREC)*.
- Smith-Stark, T. C. 1994. Mesoamerican calques. In *Investigaciones lingüísticas en Mesoamérica*, edited by C. J. MacKay and Vázquez, V. Mexico: Universidad Nacional Autónoma de México.
- Somfai Kara, D. 2002. *Kazak*. Vol. 417. München: Lincom Europa.
- . 2003. *Kyrgyz*. Vol. 423. München: Lincom Europa.
- Stachowski, M. 1993. *Dolganischer Wortschatz, Zeszyty naukowe Uniwersytetu Jagiellonskiego*. Kraków: Nakl. Uniwersytetu Jagiellonskiego.
- . 1996. *Über das Alter der Dolganischen Sprache*. Vol. 6, *Türk dilleri arastirmalari*.
- . 1997. *Dolganische Wortbildung*. Kraków: Księgarnia Akademicka.
- . 1998. *Dolganischer Wortschatz: supplementband*. Krakow: Nakl. Uniwersytetu Jagiellonskiego.
- . 1999. *Konsonantenadaptation Russischer Lehnwoerter im Dolganischen*. Kraków: Księgarnia Akademicka.
- Stachowski, M. and Menz, A. 1998. Yakut. In *The Turkic languages*, edited by L. Johanson and Csató, E. London - New York: Routledge.
- Stapert, E.L. 2009. Universals in language or cognition? Evidence from English language acquisition and from Pirahã. In *Language complexity as an evolving variable*, edited by D. Gil and Sampson, D. and Trudgill, P. Oxford: Oxford University Press.
- . Database of annotated Dolgan texts.
- Steele, S. 1978. Word order variation: a typological study. In *Universals of human language IV: Syntax*, edited by J. H. Greenberg, Ferguson, C.A and

- Moravcsik, E.A. Stanford: Stanford University Press.
- Stern, D. 2005. Taimyr Pidgin Russian (Govorka). *Russian linguistics* 29 (3):289-318.
- . 2009. The Taimyr pidgin Russian morphology enigma. *International journal of bilingualism* 13 (3):378-395.
- Stolz, C. and Stolz, T. 1996. Funktionswortentlehnung in Mesoamerika. Spanisch-amerindischer Sprachkontakt (Hispanoindiana II). *Sprachtypologie und Universalienforschung* 49:86-123.
- Sunik, O.P. 1962. *Glagol v tunguso-man'chzhurskikh yazykakh* [The verb in Tungus-Manchur languages]. Moskva Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- Tenishev, E.R. 2001. *Sravnitel'no-istoricheskaya grammatika tyurkskikh yazykov: leksika* [A comparative-historical grammar of the Turkic languages: lexicon]. Moscow: Nauka.
- . 2006. *Sravnitel'no-istoricheskiy slovar' tyurksikh yazykov: pratjurkskiy yazyk-osnova; kartina mira praturkskogo etnosa po dannym yazyka* [A comparative-historical dictionary of the Turkic languages]. Moscow: Nauka.
- Teresh'enko, N.M. 1979. *Nganasanskiy yazyk* [The Nganasan language]. Leningrad: Nauka, Leningradskoye otdeleniye.
- Thomason, S. G. 2001. Contact-induced typological change. In *Language typology and language universals. An international handbook*, edited by M. Haspelmath, E. König, E., Oesterreicher, W. and Raibe, W. Berlin: Walter de Gruyter.
- . 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
- . 2003. Contact as a source of language change. In *The handbook of historical linguistics*, edited by B. D. Joseph and Janda, R.D. Malden: Blackwell.
- . 2010. Contact explanations in linguistics. In *The handbook of language contact*, edited by R. Hickey. Oxford: Wiley-Blackwell.
- Thomason, S. G., and Kaufmann, T. 1988. *Language contact, creolization, and genetic linguistics*. Berkley: University of California Press.
- Thompson, S. A. 2002. Object complements and conversation: towards a realistic account. *Studies in Language* 26 (1):125-164.
- Thompson, S.A. and Longacre, R.E. 1985. Adverbial clauses. In *Language typology and syntactic description: Vol II: Complex constructions*, edited by T. Shopen. Cambridge: Cambridge University Press.
- Thurston, W.R. 1987. Processes of change in the languages of north-western New Britain. In *Pacific linguistics*. Canberra: Australian National University.
- . 1989. How exoteric languages build a lexicon: esoterogeny in West New Britain.

- In *VICAL 1: Oceanic languages: papers from the Fifth International Conference on Austronesian Linguistics*, edited by R. Harlow and Hooper, R. Auckland: Linguistic Society of New Zealand.
- Tiersma, P.M. 1982. Local and general markedness. *Language* 58 (4):832+849.
- Traugott, E.C. and Dasher, R.B. 2005. *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Treffers-Daller, J. 1998. Variability in code-switching styles: Turkish-German code-switching patterns. In *Code-switching Worldwide*, edited by R. Jacobson. Berlin: Mouton de Gruyter.
- Troitskiy, V.A. 1987. *Khatanga: goroda i poselki Krasnoyarskogo Kraya [Khatanga: cities and villages of the Krasnoyarsk Region]*. Krasnoyarsk: Knizhnoye Izdatel'stvo.
- Ubryatova, E. I. 1982. *Grammatika sovremennogo yakutskogo literaturnogo yazyka, fonetika i morfologiya [Grammar of the modern literary Yakut language, phonetics and morphology]*. Moskva: Nauka.
- . 1985. *Yazyk noril'skikh dolgan [The language of the Norilsk Dolgans]*. Novosibirsk: Nauka, Sibirskoye otdeleniye.
- Ubryatova, E. I., Alekseev, N.A., ed. 2000. *Fol'klor Dolgan [Folklore of the Dolgans]*. Vol. 19, *Pamyatniki fol'klora narodov sibiri i dal'nego vostoka*. Novosibirsk: Izdatel'stvo instituta arkheologii i etnografii SO RAN.
- Van Coetsem, F. 1995. Outlining a model of the transmission phenomenon in language contact. *Leuvense Bijdragen* (84):63-85.
- . 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Universitätsverlag C. Winter.
- Vasilievich, G.M. 1936. *Sbornik materialov po Evenkiyskomu (Tungusskomu) fol'kloru [Collection of Evenki materials]*. Leningrad: Izdatel'stvo instituta narodov severa SSSR im. P.G. Smidovicha.
- . 1966. *Istoricheskiy fol'klor evenkov [Historical folklore of the Evenks]*. Moscow-Leningrad: Nauka.
- . 1968. *Dialektologicheskiy slovar' evenkiyskogo yazyka [Dialectal dictionary of the Evenk language]*. Leningrad: Nauka.
- Ventsel, A.I. 2005. *Reindeer, rodina and reciprocity: kinship and property relations in a Siberian village*. Vol. 7. Berlin: LIT Verlag.
- Voronkin, M.S. 1999. *Dialektnaya sistema yazyka sakha, obrazovaniye, vzaimodeystviye s literaturnym yazykom i kharakteristika [The dialectal system of the Sakha language, formation, interaction with literary language and characteristics]*. Novosibirsk: Nauka: Sibirskaya Izdatel'skaya firma RAN.

- Voronkin, V.M. 1995. *Dialektologicheskiy slovar' yazyka Sakha: dopolnitel'nyy tom* [Dialectal dictionary of the Sakha language: supplement]. Novosibirsk: Nauka.
- Weinreich, M. 1945. Der YIVO un di problemen fun undzer tsayt. *YIVO Bleter* 25 (1): 3-18.
- Weinreich, U. 1953. *Languages in contact, findings and problems*. New York: Publications of the linguistics circle of New York 1.
- Whitten, M.C. et. al. in preparation.
- Wilkins, D. 1981. Towards a Theory of Semantic Change, Linguistics, SGS, ANU.
- . 1996. Natural tendencies of semantic change and the search for cognates. In *The comparative method reviewed: Regularity and Irregularity in Language Change*, edited by M. Durie and Ross, M. New York: Oxford University Press.
- Winford, D. 2003. *An introduction to contact linguistics*. Oxford: Blackwell.
- . 2010. Contact and borrowing. In *The handbook of language contact*, edited by R. Hickey. Malden, Oxford: Wiley-Blackwell.
- Wray, A. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, A. and Grace, G.W. 2005. The consequences of talking to strangers: evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* (117):543-578.
- Wurm, S.A. 1996. Siberia 1650-1950 ethnic and linguistic changes. In *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*, edited by S. A. a. M. P. Wurm, Tryon, D.T. New York: Mouton de Gruyter.
- Yngve, V.H. 1961. The depth hypothesis. In *Structure of language and its mathematical aspects*, edited by R. Jakobson: Providence.
- Yuasa, E. and Sadock, J.M. 2002. Pseudo-subordination: A mismatch between syntax and semantics. *Journal of Linguistics* 38 (1):87-111.
- Ziker, J. P. 1998. *Kinship, exchange, and ethnicity among the Dolgan and Nganasan of northern Siberia*. Ann Arbor: UMI.
- . 2002. *Peoples of the tundra : northern Siberians in the post-communist transition*. Prospect Heights, Ill.: Waveland Press.

Websites:

http://www.perepis-2010.ru/results_of_the_census/results-inform.php.

<http://www.taimyr24.ru>.

Software:

ELAN: psycholinguistics, Max Planck Institute for, the language archive. Available from <http://tla.mpi.nl/tools/tla-tools/elan>.

R: lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-33.

R: a language and environment for statistical computing, version 2.10.1. R Foundation for Statistical Computing, Vienna.

SUMMARY

This dissertation is a study of contact-induced linguistic change in Dolgan and explores the role linguistic data can play in the reconstruction of a people's (pre)history. While the study is primarily linguistic, its second main goal is to interpret the linguistic results in an interdisciplinary context, using insights from history, ethnography and population genetics. Thus, it provides an example of the innovative ways data from different disciplines can be combined to gain a deeper understanding of a people's past and identity.

Dolgan is a Turkic language spoken on the Taimyr Peninsula in northern Siberia and in the Anabar region of Yakutia. The history of the Dolgan people is characterised by contact between different populations, in particular between the Turkic Sakha and the Tungusic Evenks. In the literature, the Dolgans are often described as a mixed people, with both Turkic and Tungusic influences. However, it is not clear whether they were Turkic people who adopted Tungusic cultural features and a Tungusic name (Dolgan), or that they were mainly Tungusic groups who adopted a Turkic language. By combining insights from different disciplines, this study can shed new light on these issues.

Chapter 1 sets out the aims of the thesis, provides the geographical and sociolinguistic details of the field sites and launches the methodological framework used to identify contact-induced change.

Chapter 2 provides a detailed picture of the geographical, historical and ethnographic characteristics of the Dolgan people and introduces the linguistic situation on the Taimyr. It illustrates the complex history of the Dolgan people, and addresses the role politics can play in the appearance and disappearance of a people on the ethnographic map. It is shown that state ethnographers did not restrict themselves to a description of the indigenous people of Siberia, but also shaped and manipulated their identity according to their personal and political goals. This subjectivity complicates a reconstruction of the 'facts' in the ethnic history of the Dolgans, but this problem can partly be solved by the more objective data acquired through DNA-analysis. Analysis of the mtDNA (the maternal line) shows that women were very mobile on the Taimyr Peninsula, and that they often married into ethnic groups different from their own. Analysis of the Y-chromosome shows admixture between ethnic groups in the paternal line, and in particular with the Tungusic Evenks. The genetic profile of the Dolgans reveals

that one third, and potentially two thirds (depending on the analysis) of the Dolgan population has Tungusic ancestors, confirming the hypothesis of intense Tungusic contact.

Chapter 3 gives an overview of the field of contact linguistics, and introduces concepts from theories of language contact that are essential for the analysis of contact-induced change in the chapters to follow. In absence of a single all-encompassing model of contact-induced, in this study an eclectic approach is adopted, using concepts from several theories that proved useful 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 differences between Dolgan and Sakha. After an introduction to the analytical framework that is employed for the analysis of lexical change, six types of difference are analysed in both a quantitative and a qualitative way. For the quantitative analysis, the proportion of differences between Dolgan and Sakha is determined for 24 semantic fields. For the qualitative analysis, the focus shifts towards the six types of difference themselves. An investigation of their relative frequency shows that the most common type of difference between Dolgan and Sakha is semantic change and a detailed analysis of this type of difference reveals important changes in the semantic structure of kinship terminology as well as the semantic field 'the body'. The lexical form of these words corresponds to the one used in Sakha, but the semantic structure matches the one of Evenki. This leads to the conclusion that most probably this change took place by groups of Evenks who shifted to Dolgan, but kept their traditional social structure. Finally replacement is discussed, analysing copies from both Evenki and Russian.

Chapter 5 discusses morphological 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.

The next difference concerns word order patterns. Chapter 7 shows 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 text corpus of Dolgan reveals a high percentage of sentences with SVO order. While a language-internal explanation for this difference cannot be excluded, a more plausible explanation seems to be the increasing social and linguistic dominance of Russian in the Dolgan community, 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. 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. It is shown that Turkic Sakha and Tungusic Evenks are the primary ancestors of the current Dolgan population. Contact between these groups may go back as far as the 17th century, but the crucial period was the end of the 18th and the 19th centuries, when several groups of Sakha and Evenks settled along the Khatanga Trading Way and engaged in the developing trade with the Russians. Integrating anyone who participated in this new way of life, a new community emerged that adopted Sakha as its main language of communication, which had already been in use as a *lingua franca*. The language that developed in this setting later became known as the Dolgan language. It is based on Sakha, but shows lexical and structural influences from Evenki and Russian. In the second half of the 20th century, Dolgan became the official ethnonym of this diverse and dynamic group in arctic Siberia.

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

SAMENVATTING

Dit proefschrift is een studie naar taalverandering door taalcontact in het Dolgan, en onderzoekt de rol van linguïstische informatie bij de reconstructie van de (pre)historie van een bevolkingsgroep. Het onderzoek is in eerste instantie linguïstisch van aard, maar een tweede belangrijk doel is de interpretatie van de linguïstische resultaten in een interdisciplinaire context, waarbij gebruikt wordt gemaakt van inzichten uit de geschiedenis, etnografie, en populatiegenetica. Hiermee levert het een voorbeeld van de vernieuwende manier waarop data van verschillende disciplines gecombineerd kunnen worden om diepgaander inzicht te verwerven in de geschiedenis en identiteit van een bevolkingsgroep.

Dolgan is een Turkse taal die gesproken wordt op het schiereiland Taimyr in het uiterste noorden van Siberië en in het Anabar district in Jakoetië. De geschiedenis van de Dolgans wordt gekenmerkt door veelvuldige contacten met verschillende etnische groepen, in het bijzonder met de Turkse Sacha en de Toengoesische Evenken. In de literatuur worden de Dolgans beschreven als een gemengd volk met zowel Turkse als Toengoesische invloeden, maar er is onenigheid over hun oorsprong. Waren zij een Turks volk met Toengoesische culturele kenmerken en een Toengoesische naam (Dolgan), of waren zij een Toengoesisch volk dat een Turkse taal heeft aangenomen? De multidisciplinaire benadering die toegepast wordt in dit onderzoek werpt een nieuw licht op deze kwesties.

Hoofdstuk 1 presenteert de doelstellingen van de dissertatie en levert de geografische en sociolinguïstische informatie over de veldwerklocaties. Ook wordt in dit hoofdstuk het methodologische kader gelanceerd dat wordt toegepast bij de identificatie van taalverandering door taalcontact.

Hoofdstuk 2 geeft een gedetailleerd overzicht van de geografie, de geschiedenis en de etnografische achtergrond van de Dolgans zoals beschreven in de literatuur, en bespreekt de linguïstische situatie op de Taimyr. Het illustreert de complexiteit van hun geschiedenis, en bespreekt de rol die politiek kan spelen bij het verschijnen en verdwijnen van bevolkingsgroepen op de etnografische kaart. Het blijkt dat de staatsetnografen in de Sovjet Unie zich niet beperkten tot het leveren van een objectieve beschrijving van de Siberische volkeren, maar dat zij ook actief hun identiteit vormden en manipuleerden, gestuurd door hun politieke en persoonlijke doelen. Deze subjectiviteit bemoeilijkt een reconstructie van de

‘feiten’ in de etnische geschiedenis van de Dolgans, maar dit probleem kan gedeeltelijk opgelost worden door gebruik te maken van objectievere gegevens verkregen door DNA-analyse. Uit analyse van het mitochondriale DNA (dat gebruikt wordt voor reconstructie van de vrouwelijke lijn) blijkt dat vrouwen zeer mobiel waren op de Taimyr en dat zij regelmatig trouwden met mannen uit andere etnische groepen. De analyse van het Y-chromosoom laat zien dat er ook in de mannelijke lijn vermenging heeft plaatsgevonden. Eén derde, en mogelijk zelfs twee derden (afhankelijk van de analyse) van de Dolgans heeft Toengoesische voorouders, wat de hypothese bevestigt dat er intens contact bestond met deze bevolkingsgroep.

De theoretische achtergrond voor de linguïstische analyse wordt geleverd in Hoofdstuk 3. Het geeft een overzicht van het onderzoeksveld van taalcontact en taalverandering, en introduceert concepten die essentieel zijn voor de analyse van de taaldata in de volgende hoofdstukken. Aangezien er nog geen allesomvattend model bestaat voor de analyse van taalcontactfenomenen, wordt in dit proefschrift een eclectische benadering gehanteerd waarbij concepten uit verschillende theorieën worden gebruikt die het meest zinvol zijn voor de analyse van de Dolgan data. Het hoofdstuk sluit af met een discussie van de functie van theorieën voor de studie van taalverandering door taalcontact.

In hoofdstuk 4 worden lexicale verschillen tussen Dolgan en Sakha behandeld. Na de introductie van het toegepaste analytische kader worden zes typen lexicale verandering geanalyseerd, op zowel een kwalitatieve als een kwantitatieve manier. In de kwantitatieve analyse wordt het percentage lexicale verschillen tussen Dolgan en Sacha bepaald voor 24 semantische velden. In de kwalitatieve analyse worden de zes typen lexicale verandering onder de loep genomen. Uit een onderzoek naar hun relatieve frequentie blijkt dat semantische verandering het meest voorkomende lexicale verschil is tussen Dolgan en Sacha. Een gedetailleerde analyse van deze semantische veranderingen onthult interessante verschuivingen in de semantische structuur van termen op het gebied van verwantschapsrelaties en in het semantische veld ‘het lichaam’. De lexicale vorm van deze woorden in Dolgan is hetzelfde als in Sacha, maar de semantische structuur ervan komt overeen met die in Evenki. Geconcludeerd wordt dat dit type semantische verschuiving hoogstwaarschijnlijk veroorzaakt werd door Evenken die overgingen tot het spreken van Dolgan, maar hun traditionele sociale structuur behielden. Tot slot worden leenwoorden in het Dolgan besproken, zowel uit het Evenki als uit het Russisch.

In Hoofdstuk 5 staan morfologische verschillen tussen Dolgan en Sacha centraal, met name in de vervoeging van het hulpwerkwoord *e-* 'zijn' en in de verbuiging van zwakke naamwoordstammen. In Sacha hebben deze vormen een onregelmatig paradigma, maar in Dolgan zijn er regelmatige varianten ontstaan. Hoewel er uitdrukkelijk ruimte wordt opengelaten voor taal-interne motivaties voor dit verschil, wordt ook de mogelijkheid gesuggereerd dat deze regularisatie werd gestimuleerd door Evenken die Dolgan leerden als hun tweede taal.

Het volgende verschil betreft het participium habitualis. Uit analyse van het gebruik van deze vorm in tekst corpora van Dolgan en Sakha blijkt een significant verschil tussen de twee talen in zowel de morfosyntactische eigenschappen van dit participium als in de gebruiksfrequentie. In tegenstelling tot Sacha, waar het participium gebruikt wordt met een nominale en een verbale functie, komt in Dolgan de nominale functie niet voor. Het verbale gebruik van het participium daarentegen heeft een veel hogere frequentie dan in Sacha. Hoewel meer comparatief onderzoek gedaan moet worden om deze hypothese te bevestigen, is het opvallend dat de manier waarop het participium in Dolgan wordt gebruikt meer lijkt op de morfosyntactische kenmerken van de habitualis in Toengoesische talen dan het gebruik in Sacha.

In hoofdstuk 7 komen verschillen in woordvolgorde aan de orde, en het blijkt dat het Dolgan veel flexibeler is op dit gebied dan Sacha. In tegenstelling tot de strikte SOV woordvolgorde die wordt toegepast in de meeste Turkse en Toengoesische talen, wordt in het gesproken tekst corpus van Dolgan een hoog percentage zinnen aangetroffen met woordvolgorde SVO. Ook hier kan een taal-interne verklaring voor deze ontwikkeling niet worden uitgesloten, maar contact met het Russisch lijkt een plausibelere verklaring. De sociale en linguïstische dominantie van het Russisch in de Dolgan gemeenschap is groot en neemt in rap tempo toe, en in het Russisch is SVO de ongemarkeerde woordvolgorde.

In Hoofdstuk 8 worden tot slot verschillen in de vorming van samengestelde zinnen besproken. Deze verschillen zijn zeer divers, en hoewel sommige het gevolg zouden kunnen zijn van contact met Evenki, lijkt de meerderheid te zijn ontstaan door de groeiende linguïstische dominantie van het Russisch. Dit hoofdstuk laat bijzonder duidelijk de complexiteit zien van de relatie tussen sociale factoren en hun linguïstische gevolgen in een taalgemeenschap, en benadrukt daarmee het belang van multicausaliteit voor verklaringen van taalverandering door taalcontact.

In hoofdstuk 9 volgt een gedetailleerde bespreking van de deelconclusies uit de voorafgaande hoofdstukken, waarbij de individuele taalveranderingen als een geheel worden beschouwd en zij worden ingebed in de historische, etnografische en genetische context beschreven in hoofdstuk 2. Het blijkt dat zowel de Turkse Sacha als de Toengoesische Evenken tot de voornaamste voorvaders van de Dolgans gerekend moeten worden. Contact tussen deze groepen bestond waarschijnlijk al in de 17^e eeuw, maar de cruciale periode voor de vorming van de Dolgan gemeenschap was het eind van de 18^e en de 19^e eeuw, toen verschillende groepen Sacha en Evenken handel begonnen te voeren met de Russen en zich vestigden langs de Chatanga handelsroute. Er ontstond een nieuwe gemeenschap, waarin iedereen die deelnam aan deze nieuwe levenswijze kon integreren en waarin etnische grenzen vervaagden. De Sacha taal, die in die tijd al diende als *lingua franca*, werd het primaire communicatiemiddel. De taal die zich hieruit ontwikkelde is wat vandaag bekend staat als Dolgan. Het is gebaseerd op het Sacha, maar heeft ook lexicale en structurele invloeden van het Evenki en van het Russisch. In de tweede helft van de 20^e eeuw werd de naam Dolgan officieel erkend als etnoniem voor deze diverse en dynamische bevolkingsgroep in arctisch Siberië.

Hoofdstuk 10 sluit af met een korte conclusie en suggesties voor toekomstig onderzoek.

КРАТКОЕ СОДЕРЖАНИЕ

В настоящей диссертации рассматриваются изменения в долганском языке, обусловленные языковыми контактами с другими языками, а также изучается вопрос о роли лингвистических данных в реконструкции предыстории долганского народа. Исследование в первую очередь лингвистическое, однако его целью также является интерпретация лингвистических данных в междисциплинарном контексте, в частности приводятся данные из истории, этнографии и популяционной генетики. Таким образом, работа представляет собой пример новаторского исследования, объединяющего данные разных дисциплин с целью более глубокого понимания этнической идентичности и истории народа долган.

Долганский язык принадлежит тюркской семье языков. Он распространен на территории полуострова Таймыр (Таймырский долгано-ненецкий округ) и в Анабарском улусе республики Якутия. В разные периоды своей истории долганы имели контакты с разными народами, особенно тесные – с якутами и эвенками. Поэтому в литературе долган часто называют смешанным народом, сформировавшимся под тюркским и тунгусским влиянием. Однако, неясно, являются ли они изначально тюркским народом, который приобрел культурные особенности тунгусов и получил тунгусское название («долганы»), или же это тунгусский народ, перешедший на тюркский язык. Объединяя данные разных дисциплин, настоящее исследование позволяет пролить свет на эти вопросы.

В Главе 1 формулируются цели настоящей диссертации, приводится географическая и социолингвистическая информация о местах, где проводилось полевое исследование, и определяются методологические рамки исследования для идентификации контактных изменений в языке.

В Главе 2 приводится детальная географическая, историческая и этнографическая характеристика долган и описывается лингвистическая ситуация на Таймыре. В этой главе рассказывается о сложной истории долганского народа, а также рассматривается вопрос о роли, которую может сыграть национальная политика в появлении или исчезновении народа на этнографической карте. Показывается, что в процессе изучения долган этнографы не только описывали этот коренной народ Сибири, но и формировали идентичность долган в соответствии с определенными

личными и политическими целями. Эта субъективность усложняет установление действительных «фактов» в этнической истории долган. Однако эта проблема может быть частично решена с учетом более объективных данных анализа ДНК. Анализ митохондриальной ДНК (происхождение по материнской линии) показывает, что женщины много перемещались по Таймырскому полуострову и часто выходили замуж за мужчин других этнических групп. Анализ данных Y-хромосомы показывает, что происхождение по отцовской линии также смешанное. Генетический анализ выявил, что, по крайней мере одна треть, а возможно и две трети (в зависимости от типа анализа) долганской популяции генетически связаны с тунгусскими популяциями, что подтверждает гипотезу об интенсивном контакте с тунгусскими народами.

Глава 3 содержит обзор литературы по контактной лингвистике. В этой главе вводятся основные понятия из теорий языковых контактов, которые используются в последующих главах. В отсутствие единой и всеобъемлющей модели изменений, вызванных языковыми контактами, в настоящем исследовании был принят смешанный подход, использующий понятия из нескольких теорий, наиболее подходящие для анализа долганского материала. Эта глава завершается обсуждением роли теории языковых контактов в изучении языковых изменений.

В главе 4 рассматриваются лексические различия между долганским языком и якутским (Саха). В начале главы описываются методы анализа лексических изменений, далее рассматриваются шесть типов лексических различий, которые анализируются как с точки зрения частотности, так и с точки зрения семантики. Для анализа частотности рассматриваемая лексика в долганском и в якутском распределена по 24 семантическим полям. В фокусе семантического анализа находятся шесть типов лексических различий, упомянутые выше. В ходе анализа относительной частотности каждого из этих различий обнаружилось, что наиболее частым различием в лексике долганского и якутского является семантический переход. Этот тип лексических различий рассматривается более подробно, выявляются важные изменения в семантической структуре терминов родства, а также в семантическом поле «части тела». По форме эти лексемы оказываются одинаковыми в якутском и долганском языке, однако семантическая структура этих полей в долганском систематически совпадает со структурой в эвенкийском языке. На основании этих данных можно сделать вывод о том,

что с наибольшей вероятностью это семантическое изменение имело место среди эвенков, которые перешли на долганский язык, но сохранили свою традиционную социальную структуру. В заключительной части главы обсуждается феномен «лексического замещения» и анализируются лексические заимствования из эвенкийского и русского.

В главе 5 обсуждаются морфологические различия между якутским и долганским языками, в частности в спряжении вспомогательного глагола *е-* 'быть' и в склонении нестабильных существительных. В якутском языке обе парадигмы нерегулярны, в то время как в долганском появились и регулярные варианты. Автор не исключает возможности развития этой разницы под влиянием внутренних изменений в долганском, однако предлагает и другое объяснение: этот процесс мог быть ускорен эвенками, которые пользовались долганским как вторым языком.

В главе 6 рассматривается причастие со значением хабитуальности. Анализ морфосинтаксических свойств этого причастия, а также его частотности показывает, что оно по-разному используется в долганском и якутском. В отличие от якутского языка, где это причастие может использоваться как существительное, в долганском оно используется только в предикативной функции и является гораздо более частотной формой. Возможно, использование этих причастных форм в долганском изменилось таким образом, что с точки зрения морфосинтаксической дистрибуции они приблизились к соответствующим формам в тунгусо-маньчжурских языках.

Следующее рассматриваемое различие проявляется в порядке слов. В Главе 7 показывается, что в долганском языке порядок слов гораздо более свободный, чем в якутском. В отличие от большинства тюркских и тунгусских языков, в которых наблюдается строгий порядок слов SOV, в долганском, как показывают данные корпуса разговорной долганской речи, предложения с порядком слов SVO встречаются значительно чаще. Хотя причиной этого изменения также могут быть внутриязыковые изменения, и это объяснение нельзя исключать из рассмотрения, более вероятной причиной изменения порядка слов в долганском автор считает возрастающее доминирование у долган русского языка, в котором порядок слов SVO является базовым.

В главе 8 анализируются различия в стратегиях оформления сложного предложения. Вероятно, некоторые из них возникли в результате контакта с эвенкийским языком, однако большая их часть развилась относительно

недавно под влиянием русского языка и в результате постепенной утраты долганами долганского языка. В данной главе сочетается описание релевантных социальных факторов и разнообразных лингвистических последствий, и таким образом подчеркивается важная роль изучения разных факторов в объяснении языковых изменений, вызванных контактами.

В главе 9 подробно обсуждаются выводы, полученные в предыдущих главах. Лингвистические результаты помещаются в исторический, этнографический и генетический контекст, описанный в главе 2, что позволяет проследить все изменения как компоненты единого процесса. Автор показывает, что предками современных долган были преимущественно две популяции – якуты и эвенки. Контакты между этими народами можно проследить начиная с XVII в., но наиболее важным периодом стал конец XVIII в. и XIX в., когда несколько групп якутов и эвенков поселились вдоль Хатангского тракта, где активно развивалась торговля с русскими. Таким образом, росла и развивалась новая общность, языком общения которой стал якутский, являющийся к тому времени *lingua franca* в этом регионе. Язык, который возник в этих условиях, позже стал известен как долганский. В основе его находится якутский язык, однако русский и эвенкийский повлияли на него как структурно, так и лексически. Во второй половине XX в. название «долганы» стало официальным этнонимом для этой разнообразной по своему составу общности в арктической Сибири.

Глава 10 содержит заключительные выводы диссертации и обрисовывает перспективу для дальнейшего исследования.

CURRICULUM VITAE

Eugénie Stapert was born in 1981 in The Hague in the Netherlands. She studied General Linguistics, Russian Language and Cultural Anthropology at the University of Amsterdam, where she received her M.A. in General Linguistics with distinction in 2005. Starting in 2006 she joined the University of Manchester to participate in a linguistic research project, and in 2008 she was accepted as a doctoral student at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. Here she conducted her PhD research as a member of the Max Planck Research Group for Comparative Population Linguistics.