

UC Berkeley
Dissertations, Department of Linguistics

Title

Linguistic Purism, Language Shift, and Contact-induced Change in Tatar

Permalink

<https://escholarship.org/uc/item/3x61t12t>

Author

Wertheim, Suzanne

Publication Date

2003

Linguistic purism, language shift, and contact-induced change in Tatar

by

Suzanne Wertheim

A.B. (Duke University) 1990

M.A. (University of California, Berkeley) 1997

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Leanne Hinton, Chair

Professor Andrew Garrett

Professor Penelope Eckert

Professor Alexei Yurchak

Spring 2003

Linguistic purism, language shift, and contact-induced change in Tatar

Copyright 2003

by

Suzanne Wertheim

Abstract

Linguistic purism, language shift, and contact-induced change in Tatar

by

Suzanne Wertheim

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Leanne Hinton, Chair

This dissertation examines both the causes and symptoms of language shift using the specific case of Tatar in present-day Tatarstan.

In Chapter One, Tatar is shown to be an example of potentially reversible “gradual language death,” where Russian is encroaching on Tatar functional domains. Bilingual Tatar speakers have a range of styles that can be organized according to level of language mixing, with pure Tatar on one end, pure Russian on the other end, and “mixed” styles in between. Members of the “Tatar Social Club”, young Tatar-Russian bilinguals, are carving out a new Tatar functional domain in an overwhelmingly Russian urban setting.

Chapter Two shows how the language ideology that I call the “Tatar discourse of purity” is a significant part of the construction of the post-Soviet Tatar identity. “Pure” Tatar is implicitly defined as Tatar without Russian influence; influences from languages not perceived as threats can be interpreted as congruent with the post-Soviet Tatar identity. This language ideology is expressed through purification movements commonly found in post-colonial contexts, most notably through orthographic and lexical reform.

In Chapter Three I examine one mixed style of Tatar where speakers unconsciously use Russian words in otherwise entirely Tatar discourse – a sign of language attrition. These

Russian words are all metalinguistic, and structure and comment on discourse. Many of them retain their Russian grammatical requirements, and this Tatar style can then be analyzed as a “composite” of Tatar and Russian. This form of language mixing and the structural alterations that go along with it indicate a previously unrecognized pathway of language attrition and language influence.

Chapter Four shows that the code-mixing of discourse-pragmatic words both explains and gives motivation for lexical and structural borrowing in other sociolinguistically similar contact situations. I present a three-stage model that, while innovative, also derives from general linguistic principles, including theories of discourse markers and of grammaticalization.

In Chapter Five, I synthesize the findings and theories of the previous four chapters, and suggest the ways in which they can be applied in other avenues of linguistic inquiry.

This dissertation is for my parents,
without whose support, on every level,
none of this would have been possible.

Linguistic purism, language shift, and contact-induced change in Tatar

Table of Contents

List of Maps	iii
List of Tables	iii
List of Figures.....	iii
<i>Chapter One: Introduction</i>	1
1.1 The history of Tatars in Tatarstan.....	4
1.2 Tatar in the framework of endangered language theory.....	20
1.2.1 Language attitudes	30
1.2.2 Tatar’s functional domains	34
1.3 The “Tatar Social Club”	45
1.4 Fieldwork, the “observer’s paradox,” and data contextualization.....	56
1.4.1 Observation and audience design.....	59
1.4.2 The role of the language investigator in style shifting.....	67
1.4.3 The role of the language investigator in recording speech.....	72
1.4.4 The value of performance speech.....	76
1.4.5 Conclusions	80
<i>Chapter Two: Linguistic purism and the “de-Russification” of Tatar</i>	84
2.1 The Tatar “discourse of purity”	84
2.2 “Impure” Tatar	96
2.2.1 Phonetic interference	99
2.2.2 Spelling and translation errors	100
2.2.3 Code switching.....	101
2.3 Tatar purification movements	102
2.3.1 Orthographic reform and symbolic alphabets	106
2.3.2 Lexical reform: The return of Arabic and Persian borrowings.....	122
2.3.3 A comparison with 20 th century Turkish language reform.....	131
2.4 Conclusions	140
<i>Chapter Three: The code-mixing of Russian words in “Tatar-preferred” style</i>	144
3.1 Introduction	144
3.1.1 Data selection	151
3.1.2 Terminology	153
3.1.2a Code-mixing vs. code-switching.....	153
3.1.2b Discourse-pragmatic words	153
3.2 Varieties of Tatar performance.....	165
3.2.1 “Tatar on-stage” style.....	168
3.2.2 “Tatar-preferred” style.....	170
3.2.3 Other Tatar styles	171
3.2.4 Methods of data collection.....	174
3.2.4 Final thoughts on “purity”	177
3.3 Past function-word borrowing in Tatar	178

3.4 Present-day code-mixing of Russian words.....	181
3.4.1 Markers of discourse structure and force.....	182
3.4.1a Discourse-structure marking – subordinative.....	182
3.4.1b Discourse-structure marking – coordinative	186
3.4.1c Discourse-structure marking – contrastive	187
3.4.1d Metacommentary on text	189
3.4.2 Interactional performatives	195
3.4.3 Evaluatives.....	197
3.4.3a Positive evaluation	198
3.4.3b Negative evaluation.....	199
3.4.3c Hedges	200
3.4.3d Epistemics	204
3.4.3e Evaluations of time and degree	204
3.5 Russian words that are <i>not</i> code-mixed in Tatar-preferred style.....	210
3.6 Conclusions	212
<i>Chapter Four: The role of discourse pragmatics in lexical and structural borrowing.....</i>	<i>214</i>
4.1 A new model of lexical and structural borrowing	214
4.2 Case Study One: The influence of Slavic on Karaim.....	229
4.3 Case Study Two: Arabic and Persian influences in Tatar	252
4.3.1 Arabic discourse-pragmatic words found in present-day Tatar.....	254
4.3.2 Persian discourse-pragmatic words found in present-day Tatar.....	262
4.3.3 Arabic and Persian discourse-pragmatic words in context.....	267
4.3.4 Archaic Arabic and Persian borrowings.....	274
4.3.4a Archaic Arabic borrowings.....	275
4.3.4b Archaic Persian borrowings	282
4.3.4c Archaic Arabic and Persian grammatical borrowings	285
4.4 Case Study Three: Turkish influence on Asia Minor Greek	288
4.4.1 Turkish discourse-pragmatic borrowings in Asia Minor Greek	290
4.4.2 Other aspects of Turkish influence in Asia Minor Greek	312
4.5 Conclusions	318
<i>Chapter Five: Conclusions</i>	<i>327</i>
<i>Bibliography.....</i>	<i>344</i>
<i>Appendices.....</i>	<i>355</i>
Appendix A Transcription conventions and grammatical abbreviations	355
Appendix B. Russian discourse-pragmatic words code-mixed in Tatar	357
Appendix C. Arabic discourse-pragmatic words borrowed into Tatar.....	363
Appendix D. Persian discourse-pragmatic words borrowed into Tatar.....	369
Appendix E. Turkish discourse-pragmatic words borrowed into Asia Minor Greek	371
Appendix F. A cross-linguistic comparison of discourse-pragmatic borrowings	373

List of Maps

Map 1. <i>Tatarstan and surrounding administrative regions</i>	vii
Map 2. <i>Autonomous regions of the Russian Federation</i>	vii
Map 3. <i>Ethnic makeup of western Russian Federation republics</i>	viii

List of Tables

Table 1. <i>Tatar-specific phones</i>	99
Table 2. <i>English discourse markers found in Irish radio performance</i>	223
Table 3. <i>Hebrew discourse markers found in Israeli English conversations</i>	226
Table 4. <i>Russian discourse-pragmatic words found in present-day Uzbek</i>	227
Table 5. <i>“Reverse grammaticalization” in minority-language borrowing</i>	322

List of Figures

Figure 1. <i>Hierarchy of attributes and audience roles</i>	61
Figure 2. <i>Cline of language mixing and performance styles for Tatar bilinguals</i>	65
Figure 3. <i>Effect of changing participant roles of fieldworker on Galimä’s speech</i>	71
Figure 4. <i>Adjusted hierarchy of attributes and audience roles for a recorded speech event</i>	73
Figure 5. <i>‘The Latin-based Alphabet’</i>	106
Figure 6. <i>Tatar alphabet a-l from Latin Grafikası Nigezendä Älifba</i>	107
Figure 7. <i>Tatar alphabet m-z from Latin Grafikası Nigezendä Älifba</i>	108
Figure 8. <i>Graphics from an article on runic writing (Kurbatov 2000)</i>	113
Figure 9. <i>Rune-style Tatar font (Miras 2000)</i>	114
Figure 10. <i>The religious newspaper Iman ‘Belief’</i>	115
Figure 11. <i>Chistoi (Чистой) bälzäm</i>	117
Figure 12. <i>Gabdulla Tukay’s poem I Tugan Tel, calligraphed and illustrated by Näjip Näkekash</i> . 120	
Figure 13. <i>Cline of language mixing and performance styles for Tatar bilinguals</i>	165
Figure 14. <i>Results of increase in level of verbal hygiene (synchronic style-shifting)</i>	218
Figure 15. <i>Changes in composition of the borrowed lexicon (diachronic model)</i>	219
Figure 16. <i>Reassignment of styles by next Generation of speakers</i>	220
Figure 17. <i>Repeated borrowing of ‘and’</i>	318
Figure 18. <i>Repeated borrowing of ‘of course’</i>	319

Acknowledgments

As is true with most major projects, this dissertation would not have been possible without significant institutional and individual support. Financial support for my fieldwork was provided by an IREX *Individual Advanced Research Opportunities* fellowship and by an Academy for Educational Development *David L. Boren* fellowship; a follow-up fieldwork visit was financed by a *Dissertation Field Research* fellowship from the Berkeley Program in Soviet and Post-Soviet Studies. The *Chancellor's Dissertation Fellowship* from the University of California, Berkeley and a *Summer Research* fellowship from the Berkeley Program in Soviet and Post-Soviet Studies allowed me to work full time on the writing of the dissertation.

Individual support, while less easily quantifiable than institutional financial support, was also crucial. First and foremost I am grateful to the members of my dissertation committee, each of whom contributed something different to the project. Leanne Hinton's seminar in Endangered Languages sowed the first seeds for the dissertation, and her broad knowledge of literature on endangered language theory, enthusiastic support, and always constructive criticism were helpful the whole way. Andrew Garrett's intellectual rigor and demands pushed me further than I had dared to go (thanks, Andrew!), and he is a role model both as a scholar and as a mentor. Alexei Yurchak's grounding in the anthropology of post-Soviet transition studies and wide knowledge of both cultural and linguistic anthropology helped me contextualize my work and keep the focus broad. Last but by no means least, Penny Eckert's intellectual guidance and emotional support were invaluable, particularly while I was in the field, and I'm also grateful for her assistance in purchasing the sheepskin coat that kept me (mostly) warm during the long Kazan winter.

The advice, guidance, and support of people who were not acting in an official capacity were also tremendously useful. I am particularly indebted to Helen Faller, the other linguistic anthropologist who works on Tatar: her generosity with Tatar contacts, literature suggestions, time spent on close reading of my work, and general emotional support and availability was unstinting – рəхмәт сиңа, игәзәрем! Julie Lewis, the other half of my dissertation group, also provided unstinting support, close reading, and helpful suggestions, even after she had filed her own dissertation and thus had every right to abandon me. Mary Bucholtz has also been a generous and kind mentor, always positive and encouraging, and always with a useful insight or suggestion. Eve Sweetser was the first to notice the discourse-pragmatic nature of my data, and found time in her busy schedule to make sure that I was pointed in the right direction. Éva Ágnes Csató not only graciously but also speedily sent me copies of her articles that were not available here. Various people put up with my (bizarre and disjointed, at least to them) language questions: many thanks to Karl Zimmer and Ayla Algar for help with Turkish, Maria Kotzamanidou and Andrew Garrett for help with Greek, and Mino Mo'allam for help with Persian orthography. The staff of the linguistics department, Belén Flores, Esther Weiss, and Paula Floro, have been not only models of efficiency and competence, but also helpful in so many ways over the years that I can't begin to list them. The support of friends and family was also invaluable. For example, Jane Zavisca and Jen Utrata were excellent post-Soviet sounding boards and provided a sociological perspective. Other people helped in less clearly academic ways. Stephen Howard, for one, spent a lot of time and energy attempting to provide me with a social life during the dissertation-writing process – good effort! Rebecca Pevsner and Miss X. were also mood lifters and the providers of luxurious treats. Thanks also go to Ben, Julie, and Jon for the paper.

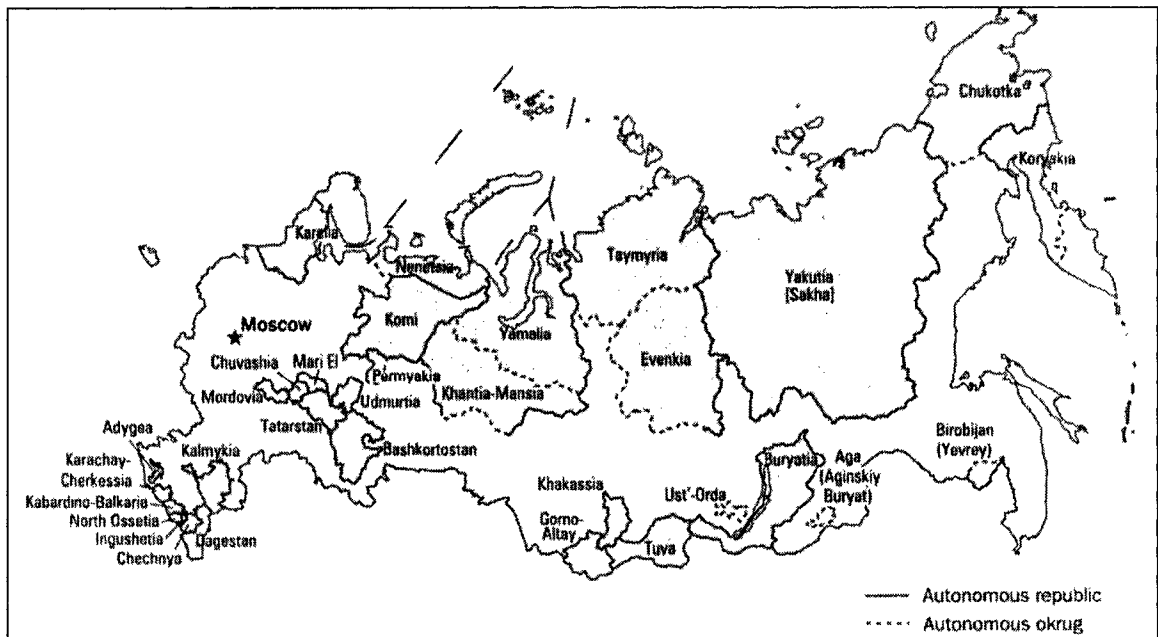
Finally, I would like to thank the many warm, hospitable, and gracious people of Kazan who helped me in so many ways. Marina and Ilya gave me not only California-style salads, but also Western-style intellectual conversations – and in English! I only hope I provided enough entertainment in return. Анна Ивановна и Владимир Александрович были как вторая семья – большое спасибо вам со всем! А спасибо друзьям в Национальном Библиотеке, которые мне так часто помогли: Гузель Солтановна, Василя Фаратовна, Наталья Викторовна, Наиля Казбековна, и моя хорошая собеседница Галия Наилевна. Ниһаят, Казандагы дусларыма, бигрәктә клуб кешеләринә, иң зур рәхмәт әйтергә теләм. Гамил, синең беренче чакырусыз бу диссертация булмас иде. Рәсимә Ханым, ярдәмеgez һәм киңәшеgez һәрвакыт файдалы булды. Айдар, аяк ярасым белән ярдәмең һәм тәмлә пылау өчен рәхмәт, чын күңелдән. Айсилу, Ләйсән(нәр), Айгөл(ләр), Алтинай, Ләйлә, Гөлнара, Руслан, Венера, Булат, Рамай, һәм төрле-төрле башка кешеләр татар теле белән дә, Казан тормышы белән дә ярдәм итте, сезгә мән рәхмәт!!!

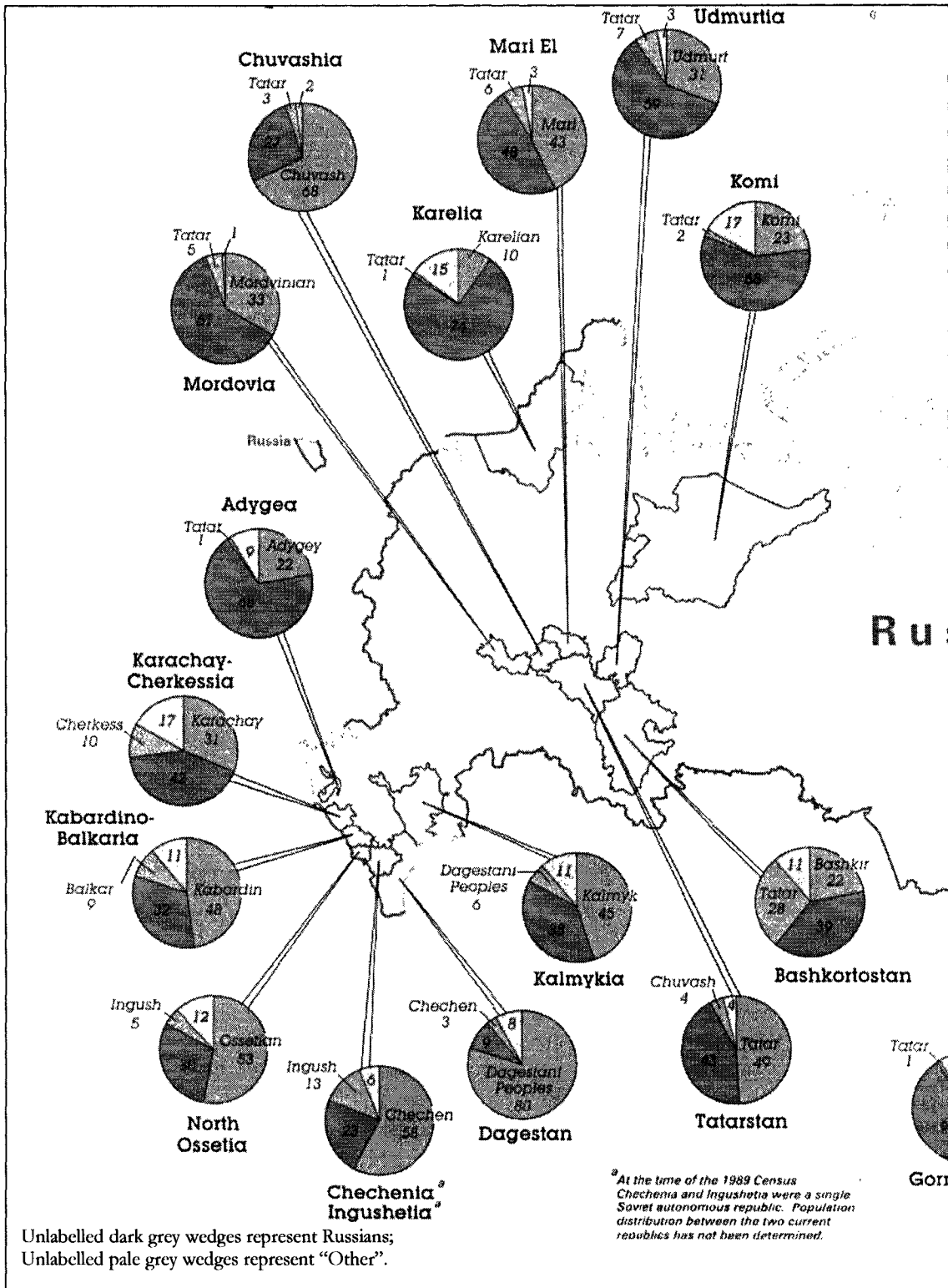
Map 1. Tatarstan and surrounding administrative regions (detail of map from the University of Texas online collection, downloaded from <http://www.lib.utexas.edu/maps/commonwealth/russiaaddivisions.jpg>)



Tatarstan

Map 2. Autonomous regions of the Russian Federation (detail of map from the University of Texas online collection, downloaded from http://www.lib.utexas.edu/maps/commonwealth/russia_auton96.jpg)





Map 3. Ethnic makeup of western Russian Federation republics (detail of map from the University of Texas online collection, downloaded from http://www.lib.utexas.edu/maps/commonwealth/russia_ethnic94.jpg)

Chapter One

Introduction

The world's languages are disappearing at an alarming rate. For example, 80% of the native languages of North America and 70% of the languages of Russia are considered to be “moribund,” that is, in their last generation of speakers (Krauss 1992: 4). In this context, the study of phenomena associated with contracting languages and language shift is increasingly important; this dissertation is part of a research agenda that addresses the struggle to retain linguistic and cultural integrity in assimilating societies. I investigate how the social, cultural, and political backgrounds of minority-language speakers affect the way that they speak, and I attempt to identify a number of the factors that contribute to the choice of some speakers of minority languages to stop transmitting their language to their children, thus causing the population of speakers to become smaller with each successive generation. In addition, I examine how language as a social practice influences language as a system, such that, for example, a minority language becomes more like the dominant language with which it is in contact. In my discussion of the causes and the symptoms of language shift, I draw on theories and findings from linguistic anthropology, sociolinguistics, historical linguistics, studies of language contact, and anthropological post-Soviet “transition studies.”

This dissertation is based on an academic year of participant-observation fieldwork among young Tatars living in Kazan, the capital of the autonomous republic of Tatarstan, which is a subject in the Russian Federation. The changing post-Soviet status of Tatars and the Tatar language, both under Russian domination for centuries, along with the current resurgence of Tatar nationalism, pan-Turkic identification and Islamic religious observance make Tatarstan an exemplary location for the study of language contact and shift, and the

relationship between language and ethnic identity. The Tatar case is an especially useful addition to the growing number of case studies of endangered languages because Tatar is at an earlier stage of language death than most other languages studied within the framework of endangered language theory.

The structure of the dissertation is as follows. In this chapter, I give the historical context of Tatars in present-day Tatarstan, I situate Tatar within the framework of endangered language theory as a case of potentially reversible “gradual language death,” I describe the young Tatars who are the focus of this dissertation, and I analyze the difficulties of contextualizing recorded and observed conversations. Section 1.1 describes the history of Tatars in the middle Volga region and examines Tatar-Russian contact from the 13th century up to the present-day. Section 1.2 gives an overview of Soviet and post-Soviet language policies and the legacy of these policies, focussing on the internal and external language attitudes towards Tatar and the asymmetry of functional domains where Tatar and Russian are used. Section 1.3 describes the locus of my fieldwork, the “Tatar Social Club,” and gives the background and orientation of the young Tatars among whom I was a participant observer. Finally, in Section 1.4 I reconsider the “observer’s paradox” and suggest methods to more systematically account for the effect of the fieldworker and recording equipment upon speech.

Chapter Two demonstrates how language ideologies can be used to link Tatar’s socio-political context and its linguistic structure. In this chapter, the Tatar “discourse of purity,” the most prominent language ideology found in post-Soviet Tatarstan, is related to purification movements on both the communal and individual levels, and contrasted with the various kinds of “impure” Tatar that are omnipresent in urban Tatarstan. “Pure” Tatar is defined in opposition to Russian, which is a selective target: while Russian influence and

interference is rejected, influences from other languages, in particular from Arabic, are not. Lexical reform, orthographic reform, and historical alphabets are used symbolically to represent cultural and political reorientation; a comparison with post-imperial 20th century Turkey shows that lexical and orthographic reforms there had similar symbolic functions.

In Chapter Three, the focus shifts to purification in individual linguistic performance and the role of style shifting in the expression of “pure” Tatar, found in a “performed” style where de-Russifying verbal hygiene is at its highest. When verbal hygiene is relaxed but Tatar remains the preferred language of conversation, speakers who believe that they are speaking “pure” Tatar unconsciously use Russian words in otherwise entirely Tatar discourse. These Russian words are all metalinguistic in nature, and organize and comment upon discourse. They retain their Russian grammatical requirements when code-mixed into Tatar, and the language produced is a “composite” that contains Russian lexical items, Russian semantic patterns, and elements of Russian morphosyntax. This code-mixing and its accompanying structural alterations are a previously undescribed pathway of language attrition and language influence.

In Chapter Four, I show that the Tatar pattern of language mixing is not an isolated case, and that adding discourse-pragmatic words as a category to the standard dichotomy of content words/system words allows for a more nuanced understanding of borrowing and structural influence. In this chapter I present a three-stage model of lexical and structural borrowing that integrates the Tatar findings with in-depth case studies and with past literature on language contact. This model, while innovative, also derives from general linguistic principles, including theories of discourse markers and grammaticalization, both of which are traditionally used in the analysis of monolingual data.

Finally, Chapter Five synthesizes the findings and theories of the previous four chapters, and suggests both future research and ways in which these findings can be applied in other avenues of linguistic inquiry.

1.1 The history of Tatars in Tatarstan

The Volga Tatars are the titular nation of the Tatar republic, known in the Soviet era as Tataria and in the post-Soviet era as Tatarstan. Tatarstan is 26,000 square miles in area (68,000 square kilometers) (Mustafin and Khuzeyev 1994: 4; Kondrashov 2000: ix), and is located in the Volga Basin at the confluence of the Volga and Kama rivers, which meet midway between Moscow and the Ural Mountains that mark the boundary between Europe and Asia.¹ Tatars are the largest ethnic minority in Russian Federation, and comprise 3.8% of its population (Kondrashov 2000: ix): this number includes not only speakers of the Kazan dialect that was the basis for standard literary Tatar, but also speakers of other dialects of Tatar, for example, Siberian Tatars and Mishar Tatars. Approximately one third of the Tatars in the Russian Federation live in Tatarstan, one fifth in neighboring Bashkortostan, and the remainder elsewhere in Russia (it is said that approximately one million Tatars live in Moscow and its environs). There are also Tatar diaspora communities found scattered across the globe, most notably in Turkey, Australia, Japan, Finland, and America (in particular in the New York and San Francisco metropolitan regions): many of these communities were established before or around the time of the Russian Revolution.

In 1993, the population of Tatarstan was 3.75 million people (Mustafin and Khuzeyev 1994: 13): approximately one quarter of the population lives in Kazan, the capital

¹ Therefore, although Tatarstan has long been considered an “East meets West” setting, it is located squarely in European Russia.

city. The most recent census results for Tatarstan come from 1989 (ibid. 18), and are as follows:

Tatars	1,765,400	(48.5%)
Russians	1,575,400	(43.3%)
Chuvash	134,200	
Others	125,000	

The ethnic groups represented in the last category are, in decreasing order: Ukrainians, Mordvins, Udmurts, Maris, Bashkirs, Belorussians, Jews, Azeris, Germans, Uzbeks, Kazakhs, Armenians, Georgians, Moldavians, and 10,000 “representatives of other nations” (ibid. 18).

As of Fall 2002, there is a new, highly politicized census ongoing in the Russian Federation.

This census is exceedingly controversial and the Radio Free Europe/Radio Liberty Tatar-Bashkir report (henceforth RFE/RL) provides updates on census problems and issues almost daily: the most controversial innovation of the census is its division of nationalities into smaller subsets. For example, Tatars will now be divided into 45 subgroups which will then be grouped into three categories: Tatars, Kreshens, and Mishars (RFE/RL October 11, 2002). Officials in Tatarstan protest that federal authorities are trying to undermine both Tatars’ cohesiveness and their status as the largest ethnic minority in the Federation through the addition of these new categories (RFE/RL September 26, 2002). Kreshens, Christian Tatars whose ancestors converted to orthodoxy in the 16th century, have been considered by census-takers for more than 70 years to be an “ethno-religious” group within the Tatar nation (RFE/RL October 10, 2002).² There have been allegations of census violations in both Tatarstan and neighboring Bashkortostan: in Tatarstan, Kreshens who wish to be

² Tatars are in general Sunni Muslims. Kreshens (also spelled ‘Kryashens’, ‘Krashens,’ and ‘Kereshens’) are a distinct minority of the population, and as is common with many sectarian groups, hold themselves separate from other Tatars. My contact with Christian Tatars was of an extremely limited nature, and they will not be addressed in this dissertation: I will be using the ethnonym “Tatar” to refer to those Tatars who are not Kreshens. For information on the Kreshen Tatars of Tatarstan, see Kefeli-Clay 2001.

counted separately from Muslim Tatars have complained that Tatarstan census takers are not registering them as Kreshens, but as Tatars, while Tatars in Bashkortostan allege that they are being registered as Bashkirs rather than Tatars, or that their census forms are being filled out in pencil, and thus will be later changed to register them as Bashkirs rather than Tatars (RFE/RL October 11, 2002).

Due to these census politics, it seems likely that the number of Tatars in the 2002 census will be lower than in the 1989 census, particularly in Bashkortostan. This would, however, contradict demographic trends and projections. For example, in comparison with the 1979 census, the number of Tatars in Tatarstan registered in 1989 rose by 123,800 people (7.5%), while the number of Russians rose by only 59,400 (3.9%) (Mustafin and Khuzyev 1994: 19). Previous censuses showed that Tatars, while always numbering more than Russians, were never the absolute majority in Tataria. In 1926, Tatars were 44.9% of Tataria's population, and Russians were 43.1%; in 1939, Tatars were 48.8% and Russians 42.9%; in 1959, Tatars 47.2% and Russian were 43.9%, and in 1970 Tatars were 49.1% and Russians 42.4% (Rorlich 1986: 172). Based on projections by Tatar demographers, it is now claimed that Tatars are presently an absolute majority in the republic with a population percentage of 51%; these claims of absolute majority have been circulating since 1995 (Graney 1996: 52).

Historians and political scientists have traditionally divided Tatar-Russian contact into three major periods: (1) the so-called "Tatar yoke," the period after Muscovite Rus' was conquered by the Golden Horde, (2) the period from the 1552 conquest of the Kazan Khanate by Ivan the Terrible to the 1917 revolution, and (3) the Soviet era (Matsuzato 2001: 47); we can now add to this a fourth period, the era of post-Soviet sovereignty. In general, Tatars have played a more prominent role in the history of the Volga Basin than their other

Turkic neighbors, and they are distinguished by their early move to agriculture, early conversion to Islam, early rise to statehood, early written culture and widespread literacy, early and strong bourgeoisie, and strong national press and intelligentsia (Graney 1999: 63).

Tatar ethnogenesis has been an area of debate since the 1940s. All theories of Tatar ethnogenesis deal with the relationship (or non-relationship) of the Volga Tatars to three sets of potential ancestors: Volga Bulgars³ (a Turkic people related to the Bulgars of what is now present-day Bulgaria), Kipchak-speaking Turks, and Mongol Turks. The Volga Bulgars are associated with the state of Great Bulgar (also known as Volga Bulgaria), which arose in the early 10th century, assimilated the indigenous Finno-Ugric population of the mid-Volga Basin, and converted to Islam in 922. The Mongol Turks, under Batu Khan, conquered and sacked the Great Bulgar state in 1236, but the language of the Kipchak-speaking Turks of the region came to predominate under their reign (Graney 1999: 53-56; Schamiloglu 1990). The Kazan Khanate, a Golden Horde successor state established in 1445 on the territory of the former Bolgar State, fell to the Russians in 1552, a conquest considered to be the beginning of the Russian Empire (Graney 1999; Rorlich 1986). Up until the 1940s, the Tatars considered their ancestral legacy to be the Golden Horde: however, a 1944 decree by the Russian Communist party criticized Tatar scholars for “misrepresenting Tatar history” through their emphasis of the Golden Horde, one of the bases of pan-Turkic solidarity and also a reminder of an unpleasant (to Russians) period of Russian history (Shnirelman 1996: 7).

“As a result of the decree, it became impossible to refer to the Golden Horde in academic or even fictional literature except in negative terms. Any sort of relationship to the people of the Golden Horde began to be treated as a disgrace, the very name ‘Tatar’ became an abusive word, and a wave of Tatarophobia ensued. Fearful of repression, Tatar intellectuals abandoned the idea of descent from the Golden Horde, thus opening a large gap in Tatar history and prehistory.

³ In Tatar and Russian the first vowel is an ‘o’: Болгар ‘Bolgar’.

To fill this gap, Tatar archaeologists, historians, linguists, and ethnographers began to make the claim that the Tatars were the descendants of the ancient Bulgars, who had maintained a state in the Middle Volga River valley from the tenth to the thirteenth centuries...” (ibid. 24).

There are at present three theories of Tatar ethnogenesis. The first is the “neo-Bulgarist” theory described in the above quote, where the Volga Tatars are posited to be the direct descendants of the Volga Bulgars. This was the official theory of ethnogenesis during the Soviet period, and some Tatars remain adherents of this theory to this day: it is posited that the neo-Bulgarist theory was used by the Soviet government to limit theoretical Tatar territorial claims to the area reigned by Volga Bulgars in lieu of the significantly larger area conquered and ruled by Batu Khan (Graney 1999: 58). The second theory, which holds very limited sway, is that Tatars are the descendants of only the Kipchak-speaking Turks of the Volga Basin (the present-day Tatar language is a Kipchak language, while present-day Chuvash is believed to be the modern descendent of the language of the Volga Bulgars (Shnirelman 1996)). The third theory holds that Volga Tatars are a combination of Volga Bulgar Turks, Kipchak Turks, and Mongol Turks, and have ethnic, cultural, and religious continuity, albeit at different levels, with all three. Along with Shnirelman (1996: 37), Matsuzato (2001: 47) links the return of this theory to the political climate:

“As Great Russian nationalism subsided, a moderate version of the theory of the Tatars’ Bulgar origin became prevalent. According to this version, the Volga Bulgars, the Golden Horde, and the Kazan Khanate were understood to have followed each other successively. That is, the Volga Bulgars were incorporated into the Golden Horde as a ruled nation but eventually assimilated and converted the ruling ethnic community to Islam...and this ethnic mixture generated the modern Tatars.”

This “synthetic” view of Tatar ethnogenesis is the theory that predominates today among both western and local Tatarstani scholars (Schalimoglu 1990, Graney 1999, Matsuzato 2001, Rorlich 1986), and is used by the post-Soviet Tatarstani government as well, who can thus “claim historical continuity for contemporary Tatarstan with both the ancient state of Volga

Bulgaria and the Kazan Khanate” (Graney 1999: 59). Volga Tatars are thus “now to be seen as one of the most ancient Turkic peoples, with a glorious history that included the organization of their own state...long before the eastern Slavs had formed one” (Shnirelman 1996: 37). And indeed the city of Kazan was founded more than 150 years earlier than Moscow: Moscow celebrated its 850 year anniversary in 1996, while Kazan will be holding its millennial celebration in the year 2005.⁴

Migration caused by territorial conflicts, drought, and plagues shifted the ruling center of the Golden Horde successor state away from the city of Bolgar in the south to lands near Kazan, and in 1445 Mahmud, son of an ousted Golden Horde khan, seized power and founded the Kazan Khanate (Rorlich 1986: 23-4).⁵ For the 107 years of its existence, the Kazan Khanate was ruled by a Kipchak-speaking aristocracy⁶ and retained elements of both Volga Bulgar and Golden Horde political and cultural structure (Graney 1999; Rorlich 1986). Despite political interference from Muscovy beginning in the mid-15th century, the Khanate remained not only independent, but economically and culturally influential beyond its borders (Rorlich 1986: 26ff). Matsuzato considers this Tatar dominance to have lasted beyond the conquest of the Khanate:

“In my opinion, along with Russians and Poles, Tatars formed one of the three greatest ‘imperial nations’ of the Russian Empire. In contrast to ethnological nations, ‘imperial nations’ means those nations who are constantly created by the population’s acceptance of the empire’s ruling religion, language, and culture. In other words, imperial nations pay little attention to genetic or physical anthropological factors of nationalities, and therefore, imperial nations are expansive by nature, constantly assimilating surrounding populations. The point

⁴ It’s not really clear how the year 1005 was chosen as the year of the founding of Kazan, particularly as archeological findings in the Kazan Kremlin point to an earlier date of settlement.

⁵ Other Golden Horde successor states included the Crimean Khanate, the Astrakhan Khanate, and the Nogai Horde (Rorlich 1986: 23).

⁶ Funerary inscriptions and religious commentaries from this period are the earliest linguistic evidence for the Kazan Tatar language (Rorlich 1986: 30).

here is the imperial nations long remained such, even after empires *per se*, such as Rzeczpospolita and the Kazan Khanate, ceased to exist” (Matsuzato 2001: 73, n. 18).

The fall of the Kazan Khanate in October of 1552 is considered to be the most momentous day in Tatar history, and it remains at the forefront of Tatar consciousness. In post-Soviet times, *Khäter Köne* ‘Memory Day’, observed on the second Sunday in October, commemorates the defense and fall of Kazan with a gathering in Kazan’s *Ploshchad’ Svobody* ‘Freedom Square’ and subsequent march to the Kazan Kremlin.⁷ Additionally, during my fieldwork some Tatars when asked, “where is your village?” would respond, “Well, originally we lived in Village X, but now we live in Village Y”; after several months, it became clear that “originally” meant not several generations ago, but before the forced resettlements of the mid-16th century.

After the conquest, some Tatars were expelled from Kazan to outside its gates; others were assimilated into the local administration and commercial functions of the Russian Empire, and are known as “service Tatars”(Graney 1999: 63). These service Tatars evolved into a powerful bourgeoisie that dominated 18th and early 19th century trade and industry, controlling the majority of large industrial enterprise, and serving as economic middlemen and “civilizing” religious emissaries in Central Asia, leading to hostility and the fear of Tatar imperialism on the part of many non-Tatar nationalities (Graney 1999, Matsuzato 2001, Rorlich 1986). Post-conquest Tatars were

“an ethno-political formation consolidated on the basis of Islamic-Turkish tradition and memories of the Kazan Khanate. The Mid-Volga Basin was a battlefield of two traditions: Russian-Slavic and Tatar Islamic. If Finno-Ugrians and other Turkish populations in this Basin did not wish to be Russified, they needed to accept another worldwide civilization, that is to be Tatarized” (Matsuzato 2001: 48).

⁷ The actual anniversary of the fall is October 15th in the Julian calendar.

The 18th century rule of Catherine II in particular was favorable to the Tatars, and the 1788 Act of Religious Tolerance passed during her reign undid some of the effects earlier anti-Muslim legislation; this act “facilitated a cultural and spiritual rebirth among Volga Tatars simultaneous with an economic development of the Tatar merchant class” (Graney 1999: 64). Tatar merchants were necessary to the Russian empire because they could facilitate trade with the Muslim states of Central Asia and the Kazakh steppes (Rorlich 1986: 42ff).

Once official bans on both printing and Muslim educational institutions were lifted, the Tatar elite established both printing houses and Tatar-oriented education at all levels, including at Kazan State University, which was founded in 1804. Up to the 19th century, books had been imported from other Islamic centers: now both secular and religious books were printed in Arabic and in Tatar at both the university publishing press and at smaller private presses (Rorlich 1986: 69-72). These publications were ubiquitous, reaching even remote rural areas, where they were sold at markets and fairs until the 1907 government ban on the village book trade, after which Tatar benevolent societies founded rural libraries (ibid. 70-73). Although there were regular attempts at Russification and conversion of Tatars between 1552 and 1917, Russian Imperial policy towards the Tatars was generally designed to

“harness their economic potential and quell potential rebellious tendencies by firmly integrating them into the Russian state administratively and politically, while at the same time allowing Tatars a significant amount of cultural and religious autonomy. Under Russian Imperial rule, the Tatar community of Kazan and the Middle Volga region flourished and became the wealthiest, most well-educated and socially advanced Muslim group in the entire Russian empire by the beginning of the 20th century” (Graney 1999: 65).

Nineteenth-century Tatar culture was particularly notable for its moderate Islamic religious and educational reform movement known as “Jadidism” (*jadid* means ‘new’ in Arabic). The Tatar intelligentsia of the time were in contact with and influenced by scholars and

philosophers in both Russia and the larger Turkic and Islamic worlds, including the Ottoman Empire, and Jadidism was an attempt to make Muslim Tatar practice both relevant to and congruent with a both a modernizing Islamic world and a Russia that was becoming increasingly Western (Graney 1999, Rorlich 1986). The leaders of the Jadid movement promoted both a pan-Turkic orientation and the usefulness of the Russian language as an entrée into the international scientific community (Rorlich 1986: 52ff). Jadid schools used a new phonetic way of teaching the Arabic alphabet for Tatar, stressed secular subjects, and made Russian a compulsory subject in addition to Arabic (ibid. 92). Shihabeddin Merjani, one of the leaders of the Jadid movement, attempted to demonstrate that learning the Russian language was not equivalent to being Russified: while promoting the acquisition of Russian-language skills, he refused to use Russian endings for Tatar family names or the Christian calendar (ibid. 52): even so, there are Tatar intellectuals today who believe that Jadidism was one of the main causes of Tatar 19th and 20th century assimilation.

Up until the mid-19th century, the Muslims of Russia were united by three common literary languages: Classical Arabic, Classical Persian, and Chagatay (a medieval Turkic literary language) (Rorlich 1986, Bennigsen and Quelquejay 1961). The proponents of Jadidism wanted to unite the Turks of Russia and Central Asia with a single educational program, a single ideology, and a single Turkic language. Ismael Bey, a Jadid leader, suggested that the common Turkic language be based on Ottoman, while other reformers preferred the literary language based on the Kazan Tatar dialect (Rorlich 1986: 57ff). In the late 19th century, Kayuum Nasıyırı began his advocacy of the development of a literary language based on the vernacular of Tatars of the Volga region, a language with minimal Arabic and Persian elements that would be accessible to all. In addition to writing grammars and dictionaries of Tatar, he translated Arabic and Ottoman literature into literary Tatar

(Rorlich 1986: 66). Tatar, both highly prestigious and prominent due to its great 19th century literature and philosophy as well as its highly productive publishing houses, began supplanting Chagatay: Mishar Tatars, Siberian Tatars and Bashkirs began using Kazan Tatar as their literary language (Bennigsen and Quelquejay 1961). Tatar intellectuals of the late 19th and early 20th centuries began promoting the expansion of Tatar's functional domains for Kazan Tatars as well; one reformer even advocated the translation of the Koran into Tatar (Rorlich 1986: 60). Such a translation is expressly forbidden by Islam; however, other elements of the Jadid movement remain in Tatar culture to this day, and Tatar Muslims practice a more liberal version of Islam than the Muslims of Central Asia or the Caucasus.

Because of the comparative wealth, prestige, and power of the Tatar nation, Tatars did not have a real incentive during the revolutionary years of 1905-1917 to seek territorial autonomy, and instead focused on increased cultural, educational and religious rights for Tatars and other Russian Muslim nationalities (Graney 1999: 65, Rorlich 1986: 119). However, in response to the political changes of 1917, Tatars changed their policy goals and decided that autonomy was key (Graney 1999: 73; Rorlich 1986: 132). In November of 1917, Tatar and Bashkir leaders called for an Idel-Ural state that would be part of a Russian Federal Republic (*Idel* is the Tatar name for the Volga river, therefore Idel-Ural marks the geographical boundaries of this theoretical republic, from the Volga river to the Ural mountains).⁸ However, the Idel-Ural republic, which would have united the 5-6 million Muslims of the region, was not to be: the Bashkir ASSR (Autonomous Soviet Socialist Republic) was created in March 1919, and the Tatar ASSR ten months later, on May 27, 1920

⁸ The Idel-Ural movement of 1917 and the post-Soviet Idel-Ural movement have different geographical boundaries: the territory proposed in 1917 is essentially the equivalent of today's Tatarstan and Bashkortostan, while modern proponents of Idel-Ural autonomy propose a larger territory that would include other Volga Basin Turkic peoples (Graney 1999; also author's notes from interviews with Idel-Ural nationalists).

(Bennigsen and Quelquejay 1961). Border-drawing reflected the Soviet “divide and conquer” policy and fear of Tatar domination of the middle Volga region: only one quarter of the Tatars in the Russian republic were included in the borders of Tataria, the Tatar ASSR, and Russian outnumbered Tatars in Bashkiria – Bashkirs were third in number in their titular republic, and remain so to this day (Graney 1999, Bennigsen and Quelquejay 1961). It has been posited that the Tatar ASSR’s lack of external borders prevented it from becoming the sixteenth union republic, and that this “humiliating status” as ASSR when Central Asian republics were granted Union Republic status consolidated Tatar national consciousness (Matsuzato 2001: 48-9). The organization of the Soviet Union was based on “ethnofederalism”;⁹ in addition to Russia there were fifteen SSRs (Soviet Socialist Republics), also known as Union Republics. These Union Republics, polities such as Estonia and Uzbekistan, had the highest possible level of sovereignty, including their own constitutions and state symbols, and the theoretical right to secede from the Soviet Union. Next were the twenty ASSRs (Autonomous Soviet Socialist Republics), mostly ethnically-based, which also had their own constitutions and state symbols, but less political autonomy; examples include the Tatar and Bashkir ASSRs. Lower on the scale of autonomy were the eight Autonomous *Oblasts* (‘regions’), also ethnically designated, and lower still were the twenty ethnically designated Autonomous *Okrugs* (‘areas’). Finally came the 57 *krais* (‘territories’) and *oblasts* of the Russian Soviet Federated Socialist Republic (RSFSR) (Graney 1999: 76). The early Soviet nationalities policy, which will be discussed further in section 1.2 below, was “national in form, socialist in content”: after first developing national cultures through the policies of

⁹ For more on the Soviet system of “ethnofederalism” (also referred to as “ethnic federalism”) see Connor 1984 and Suny 1993.

ethnofederalism, the ultimate goal would be the elimination of national differences and the creation of a universal Soviet culture through *sblizheniye* 'bringing together' and *sliyaniye* 'merger' (ibid. 77; see also Kreindler 1995 and Rorlich 1986, who dates these ideas back to the 1870s). Efforts to build this universal Soviet culture in the 1920s and 1930s included the "systematic destruction of the traditional social and cultural institutions which had served to construct and transmit the group identities of the Russian Empire's non-Russian minorities before the Soviet period" (Graney 1999: 79), and the purging of Tatar nationalist and cultural leaders (Rorlich 1986: 155). In addition to closing down independent Tatar publishing houses, Soviet leaders shut down the Tatar ASSR's Islamic religious and educational institutions, and closed off educational and religious exchanges with Turkish and Central Asian Muslims (ibid. 79).

The periods of *glasnost*' and *perestroika* of the late Soviet era meant that nationalistic arguments were heard once more in the Soviet Union of the mid-1980s, and Tatar leaders set out to obtain the status of a Union Republic (Matsuzato 2001: 49). Estonia was the first Union Republic to declare sovereignty in November of 1988. While visiting Kazan in August of 1990, Boris Yeltsin advised the Tatars to "take as much sovereignty as you can swallow" (Matsuzato 2001: 66), and Tatarstan declared itself to be a sovereign state on August 30, 1990. The Soviet Union was entirely dissolved on December 25, 1991 (Graney 1999: 80). In 1991, all but one the autonomous *oblasts* of the former Soviet Union were renamed 'republics', and the Russian Federation today is composed of 89 *subyekty* ('subjects'): 21 republics (of which Tatarstan is one), 1 autonomous *oblast*, 10 autonomous *okrugs*, 55 *oblasts*, and 6 *krais* (Walker 1996: 1). As with many other *subyekty* of the Russian Federation, there is a strong continuity between present-day elites and pre-1990 elites in Tatarstan (Matsuzato 2001: 49), most notably in the President of Tatarstan, Mintimer Shäymiyev, who was CPSU

first secretary of the Soviet Tatar ASSR at the time of the transition, and had earlier been chairman of the republic's Council of Ministers (Walker 1996: 11).

The Supreme Soviet of the Tatar ASSR adopted the “Declaration of Sovereignty of the Tatar S.S.R.” in August of 1990 under pressure from daily demonstrations and rallies organized by *TOTs* (the *Tatarskii Obshchestvennyi Tsentri* “Tatar Public Center”)¹⁰ (Graney 1999: 90), a nationalist organization founded in 1988 by prominent Tatar members of the local Communist party, including Rafael Khakimov, who is now a leading presidential advisor. This declaration of sovereignty was a compromise between warring factions in Tatarstan: declaring Tatar a state language alongside Russian was in deference to nationalists, but rather than declaring Union Republic status, the more vague concept of “sovereignty” was used, allowing for a variety of interpretations (Matsuzato 2001: 66). Nationalism in Tatarstan is not monolithic, and after the collapse of the Soviet Union, Tatar radical nationalists separated from the existing Tatar national movement (ibid. 66): in the early 1990s there were an array of Tatar nationalist organizations, most of which are significantly less popular today.

Katherine Graney, a political scientist specializing in nationalism movements whose work is the source of much of the historical material presented here, labels the nationalism of the post-Soviet Tatar government a “sub-state sovereignty” project. Her claim is that since Tatarstan's 1990 declaration of sovereignty, “all political and economic activity...has been pursued under the auspices of bringing those claims of sovereignty to life” (Graney 1999: 2), and that the political elites of Tatarstan “have sought to make their sovereignty project complete or maximal in a symbolic or discursive sense, if not in an institutional or material sense” (ibid. 8). Tatarstan does not have an established independent army, border controls, or a separate currency – although these are all consistent with state sovereignty,

¹⁰ *TOTs* is referred to in Tatar as *BTIÜ*, the *Böten Tatar İjtimagiy Üzgege* ‘The All-Tatar Social Center.’

they would be “potentially disruptive” (ibid. 3). Russia has accommodated Tatarstan and Bashkortostan’s sub-state sovereignty projects “largely because it is too weak to oppose them effectively” (ibid. 3), but this may be changing under the rule of Vladimir Putin, who has reorganized the Russian Federation into seven major regions, each governed by an appointed rather than elected Presidential Representative, and has orchestrated the recent attempts to bring Tatarstan’s constitution “in line” with the Federal constitution.

The Tatarstani government devotes considerable resources to the presentation of symbols of Tatarstan’s sovereignty while largely ignoring the Russian Federation. Russian Federation flags are nearly invisible in Kazan,¹¹ and while Russian Federation holidays are generally disregarded, Tatarstan’s Constitution Day (November 6) and Sovereignty Day (August 30) are national holidays in the republic. After the declaration of sovereignty, Tatarstan adopted a new state flag, a new state seal, and a new hymn, all of which incorporate the Tatar ethnicity in some way. A flag with a small crescent moon was considered, but was decided to be too overtly Islamic and thus potentially divisive: however, the green stripe at the top of the new Tatarstan flag (which has three horizontal stripes: a thick green stripe, a thin white stripe, and a thick red stripe) has clear Islamic connotations (ibid. 287-8). The new hymn was written by a Tatar composer, and the new state seal is overtly ethnically Tatar: at its center is the *akbars*, a white winged leopard that was a symbol of Volga Bulgaria (ibid. 288-9).¹²

¹¹ In eleven months of fieldwork I saw a total of three Russian flags in Kazan, all of which were associated with Federal offices.

¹² Kazan’s hockey and soccer teams are called *Akbars*, as is a bank and Tatarstan’s major recording label.

Here we see that ethnic Tatar identifications are allowed to “manifest themselves as ‘national’ ” (ibid. 303). Tatar nationalism is not a racist nationalism – even radical nationalists who promote complete secession from the Russian Federation present the (theoretical) autonomous Tatarstan as a multinational polity (Matsuzato 2001: 53; author’s interviews). However, Soviet ethnofederalism “fostered a ‘territorialized’ sense of nationness among the titular ethnic groups of that system” (Graney 1999: 7), and while Tatarstan’s official national ideology is one of civic multiculturalism, this multiculturalism “coexists uneasily with a competing national ideology that posits the eponymous titular ethnic group as the ‘true’ and ‘authentic’ national community”, the national community in whose name the sovereign state was proclaimed (ibid. 6-7). Both Russian activists and Western scholars assert that government resources are used to promote Tatars over other ethnic groups in the republic, that “Tatar” is made to equate with “Tatarstani,” and that the government is engaged in the process of “Tatarization” of Tatarstan while ignoring the cultural needs of the republic’s non-Tatar population. Tatarstan government officials, including the Minister of Culture, have argued that promotive Tatar policies are necessary to undo the damage done to Tatar culture by the Russification policies of the Soviet era (Graney 1999: 307ff).

“...[C]ivic discourse about the multi-ethnic Tatarstani nation is complicated by the simultaneous assertion that Tatarstan is the particular and special spiritual homeland of the Tatar people. Tatarstan’s leadership contends that these two assertions are not incompatible and that some level of identification with the ethnic Tatar nation can be reconciled within the civic Tatarstani nation.” (ibid. 309).

This Tatar “affirmative action” has included increasing the percentage of ethnic Tatars in the government (the first state council of the sovereign republic was 73% Tatar, as opposed to 58% Tatar in the state council of the 1990 Tatar ASSR), prioritizing the restoration of Tatar historical landmarks and monuments over those of other Tatarstani peoples, returning mosques to Muslim communities much faster than churches to Orthodox ones, and

renaming public spaces with ethnically Tatar names (ibid. 251, 318-320). When Tatarization is seen as potentially overly disruptive, however, the Tatarstani government has chosen a more moderate course: ethnic Tatar holidays such as Muslim *Näürriş* and *Korban* are state holidays, as is Orthodox Easter, but *Khäter Köne* 'Memory Day' is not an officially sanctioned holiday; when renaming Kazan's Lenin Street, the government chose to call it Kremlin Street rather than Söyembike Street, a name proposed by some Tatars to honor the historical figure of Söyembike, 16th century Tatar princess and important legendary symbol (ibid. 290-4, 317).

Other Tatar "affirmative action" policies have focused on the promotion of the Tatar language and the development of Tatar national education. Not only have new Tatar-language schools opened since 1990, such that the number of students in Kazan studying in Tatar-language schools rose from .02% in 1990 to 26.4% in 1996 (*Luchshikh Skhol Kazani* 1996 in Graney 1999: 324), the Tatar language was made a mandatory subject for students of all nationalities beginning in 1994 (Graney 1999: 279). Tatar was declared to be an official state language alongside Russian in the 1990 declaration of sovereignty, and this was enshrined in Tatarstan law with the 1992 "Law on the Languages of the Peoples of Tatarstan." Legislation and policy recommendations in the early 1990s suggested that certain governmental posts (including President of Tatarstan) have bilingual requirements, and offered government workers a 15% bonus in pay for demonstrating a sufficient level of bilingualism: in later years, this legislation was overturned. Additional policies that promote Tatar language and use will be examined in section 1.2 below, as will the gap between what is legislated and what exists in reality.

1.2 Tatar in the framework of endangered language theory

Campbell and Muntzel (1989) defined a basic typology of language death situations: (1) sudden death, where all of a language's speakers suddenly die or are killed (e.g., Tasmanian); (2) radical death, where as a result of severe repression and genocide, remaining speakers shift immediately to the dominant language (e.g., some languages of El Salvador); (3) gradual death, with a gradual shift to the dominant language via extensive contact (e.g., Gaelic in Scotland); and (4) bottom-to-top death, where the language remains only in ritual contexts (e.g., Mayan languages of southern Mexico that are used only in religious ceremonies). Of these four categories, gradual death – a multi-generational language shift with at least one bilingual generation but frequently more – is the most useful for studying the causes and symptoms of minority language shift and maintenance. While studies in the context of endangered language theory usually involve clearly moribund languages, where the shift to the dominant language is nearly complete (with perhaps a single older generation of fully competent speakers), I was interested in studying an earlier stage of language endangerment, one that might have a chance of reversing. Although Tatar still has approximately one million speakers in Tatarstan alone, absolute number of speakers is no guarantee of a language's health, as proven in the 20th century by the spectacularly fast reduction in the number of speakers of such languages as Breton and Navajo, to name just two. Tatar in Tatarstan can be potentially classified a language in the process of the third type of language death in the above typology: a gradual death with a multi-generational language shift involving at least one bilingual generation. The most salient evidence of this language shift found in Kazan, in addition to the proportionally low amount of Tatar heard there, is the quite common sight of “mixed” conversations in which a parent or grandparent speaks to a child or teen in Tatar, and that child or teen responds in Russian. In this section I will

present justifications for this classification, including an examination of the language attitudes found in present-day Tatarstan and the functional domains where Tatar is used, but first I will contextualize these language attitudes and domains within Soviet and post-Soviet language policies. Fifty percent of the languages of the former Soviet Union are considered to be endangered, and this number rises to seventy percent when considering only the languages found in the present-day Russian Federation (Krauss 1992: 5); I believe that Tatar should be considered one of these endangered languages.

In the previous section, I gave a brief overview of Tatar history and Tatar-Russian contact over the last 450 years; here I will focus on the Soviet and post-Soviet eras. While Tatar can be considered to have been under stress from Russian since 1552, it is the intensification of this stress during the Soviet period that seems to have triggered a shift from what is known as *stable diglossia*, where two languages co-exist and are used for different purposes, to *encroaching diglossia*, where one language loses functional domains to another, often with the ultimate result of complete language shift. However, as we will see, for the past twelve years the government of Tatarstan has been engaged in “promotive” language policies in order to combat this contraction of the Tatar language.

Governmental language policies are typically divided into (1) overt policies, which are explicit and *de jure*, and (2) covert policies, which are implicit and *de facto* (Schiffman 1996: 13). The overt and covert nationalities policies of the Soviet era privileged Russian and Russians over national languages and their speakers (for detailed analyses and case studies, see the collections of articles in Kirkwood 1989 and Kreindler 1985). Schermerhorn (1970) combined the paired concept of minority and majority populations with centripetal assimilationist forces and centrifugal separatist forces to form four possible sociological models of multi-ethnic nations. Soviet attempts to assimilate the nationalities of the Soviet

Empire fall into the fourth category of Schermerhorn's typology, where the combination of a centripetal majority and centrifugal minorities leads to forced assimilation with resistance on the part of the minorities; the centripetal forces of the Soviet Union are traditionally labelled "Russification" (Marshall 1996: 10).

Mutual centrifugal tendencies held by both the minority and majority populations can lead to diglossic situations, and therefore stable bilingualism.¹³ However, the conflict of centripetal and centrifugal forces appears to be one of the causes of the transformation of stable diglossic situations into encroaching diglossia, where one language loses ground to another. Centripetal assimilationist pressures may arise from both questions of nationalism and of nationism. A term coined by Fishman (1972), *nationism* deals with the pragmatic problems of governance, while *nationalism* deals with questions of allegiance, patriotism, and identification with the nation. Both nationistic and nationalistic government policies can affect minority language status: it can be expensive and impractical to govern in multiple languages, particularly when those languages are spoken by small numbers of people, and the ethnic "otherness" of minorities, with language a particularly salient characteristic, can threaten nationalistic perceptions of a unified country. Romaine (1995: 251) posits that "sanctioning the use of another language, and by implication the cultural values it symbolizes...is seen as a threat to the dominant group in society." Fishman (1972: 194) argues that in a multilingual nation, only the dominant language "is seen as fulfilling nationwide purposes on a permanent basis, or as being linked to the...national goals, national

¹³ The term *diglossia* was coined by Charles Ferguson (1959), who was the first to note the stability of language types that are not competing for functional domains. In Ferguson's model, the language types were high and low variants of the same language: this model was later updated by Joshua Fishman (1967) to include different languages as well.

symbols, national rituals, national holidays and national identifications that...nations...need and create.” Historically, the rise of multicultural nations united by nationalistic ideals has led to intensified pressure on minority language speakers to shift to the dominant language of the nation. Hamp (1989) found that speakers of Arvanitika Albanian, found in communities that had been in Greece for approximately 750 years, had maintained linguistic stability until the rise of 19th century Greek nationalism, at which time the level of Greek interference in their Albanian increased, and they began shifting to Greek-style syntax and phonology, or to Greek altogether. It has been argued that France was unproblematically multi-ethnic and multilingual until the 1790s, but after the French Revolution needed a “unifying national identity, expressed in part by a single national language” (Dorian 1998: 5). And Kreindler (1982: 9) claims that the use of the Russian language by Soviet leaders is analogous to its use by the Romanovs at the end of the Tsarist reign as “the cement of the Empire.”

Soviet language planning can be separated into three eras: (1) the period from 1917 to the mid-1930s, (2) the period from the late 1930s to the 1950s, and (3) the period from the 1950s up to the *perestroika* era of the mid-1980s (Kreindler 1995: 189). Russification policies, of course, date to the beginning of the Russian Empire, and Russification of Tatar education began in earnest with Christian missionary schools established in the 1860s: the statues of these schools stated, “The Russification of Muslim-Tatars can be achieved only by disseminating Russian language and education” (Rorlich 1986: 86). Lenin advocated the “national in form, socialist in content” formula of the early Soviet period, and opposed making Russian the sole state language or the compulsory language of education for the minority nationalities. Under Lenin, *narodnost*’ (nationality) languages were developed so they could be used for schools and local government, and alphabets were developed for languages that had been previously unwritten: it was believed that literacy would enable the

spread of communist theory and ideas (Kreindler 1995: 190). During this first era of Soviet language policy, language corpus planning was seen as “an integral part of state construction” (ibid. 192). In the USSR of the 1920s, the onus of bilingualism was meant to be on Russians living in non-Russian areas; for example, the preface to a 1925 Tatar grammar written for non-Tatars states that “all citizens with jobs in the state...and other organizations must know Tatar, and not just the spoken variety but also its written form” (Kreindler 1982: 9). In the 1910s and 1920s, Volga Tatars, who were beginning to lose their leadership among Russia’s Muslims, had Tatar-language textbooks for primary and secondary schools, while higher level (university) education remained in Russian only (Haarman 1985). In addition, there were 139 Tatar-language newspapers being published, making Tatars the minority *narodnost’* with the highest level of newspaper publishing (Kreindler 1995: 193).

However, after Lenin’s death, Soviet language planning moved steadily away from his advocated position of language equality (Kirkwood 1989: 2; Marshall 1996: 20). In 1938, more than 70 languages were used as the medium of instruction in the schools of the Soviet Union, but Stalin’s “revival of patriotism” placed great emphasis on both the Russian language and culture: under his rule, Russian was made a compulsory school subject and all languages of the Soviet Union using a Latin-based alphabet were moved to Cyrillic alphabets (Kreindler 1982, Kreindler 1995, Marshall 1996). It has been claimed that “[t]he greatest setback for the national languages, however, was not the re-emergence of Russian by their side, but the physical elimination of the native intelligentsia in the Stalin purges” (Kreindler

1995: 194).¹⁴ Work on the development of minority Soviet languages became life-threatening, and the only safe language development became “enrichment” from Russian: not only the replacement of native terminology with Russian equivalents, but also the introduction of Russian calques and verbal phraseology (ibid. 194).

The rise to power of Khrushchev in the 1950s marks the beginning of the third era in Soviet language planning and policy. Here we find a conflict between an overt policy of “Rossification” and a covert policy of “Russification.” While English has just one adjective, *Russian*, to describe both members of the Russian ethnicity and subjects of the Russian state, Russian has two: *ruskii* for the former, and *rossiskii* for the latter. The term “Rossification” is taken from the word *rossiskii* ‘pertaining to Russia as a state’, and is a condition of multiple allegiance, with an “appreciation of nationality and language rights, combined with political loyalty to a supranational union (USSR) of equal nationality republics” (Marshall 1996: 12). “Russification,” on the other hand, is taken from the adjective *ruskii* ‘[ethnic] Russian’, and is assimilationist in nature (ibid. 12). The overt policy of Rossification didn’t quite hide the assimilationist policy of Russification with a goal of “a high rate of Russian use among non-Russians” (ibid. 17). From the 1950’s onward, the Soviet Union had a covert Russification policy of asymmetrical “national-Russian bilingualism” (Haarman 1985), where minority nationals were meant to be bilingual, but Russian speakers were not: by 1989, 48.1% of non-Russian nationals were bilingual in Russian, but only 3.5% of Russians were bilingual in a Soviet *narodnost’* language (Marshall 1996). While overt policy had granted official status to a range of languages, the covert policy was one of monolingual Russian use for government, and quite often, education as well (Grenoble and Whaley 1998, Marshall 1996).

¹⁴ While there were purges of Tatar nationalist figures in the 1920s, the purges of the 1930s focused on cultural leaders and the intelligentsia, particularly writers.

The bias towards Russian-language education begun in the 1930s was furthered in the 1950s when a law, later made a constitution article, gave parents “the right to choose the language of instruction for their children” (Marshall 1996: 21) However, officials would bias parental decisions by not informing them that *narodnost*-language schools were an option, instead simply asking if the parents wanted their children to know Russian (a question to which the answer was usually ‘yes’) and then placing the children in Russian monolingual schools (Kreindler 1985). As a result of these overt and covert policies, 31 minority languages were phased out as the medium of education between 1961 and 1982 (Kreindler 1995, Marshall 1996); in Tatarstan many Tatar-language schools were forced to close or switch to Russian-language education in the 1970s (Graney 1999: 323).

The assimilationist policies of Russification caused resistance among speakers of many *narodnost*' languages: those languages that were associated with SSRs, particularly Union Republics, had a higher rate of language maintenance than those languages spoken by nationalities without a titular republic (Marshall 1996: 19-20, 33). However, language retention was much higher among rural speakers than among the urban intelligentsia: “As the languages lost prestige and narrowed in function, the national elite, especially in the cities, began to know Russian better than their own language” (Kreindler 1995: 197). The decline of a national language often leads to a rise in nationalist sentiment and action, and this was the case in the fall of the Soviet Union and transition to sovereignty as well: informal nationalist organizations arose around language issues, and language laws were usually the first “concrete manifestations of national self-assertiveness” in the transitional republics (ibid. 198). This is true for Tatar: the 1990 Tatarstan declaration of sovereignty raised Tatar to the level of official state language alongside Russian.

Recall that, as discussed in section 1.1 above, the official national ideology of the government of sovereign Tatarstan is one of civic multiculturalism, but that Tatars are in fact privileged over other ethnic groups. While there is a token discourse in post-Soviet Tatarstan on developing all of the languages of the republic, the vast majority of resources go towards the development of Tatar (Graney 1999: 313). The promotion of Tatar language and culture are justified as a sort of “affirmative action” necessary to counteract the harm done by Soviet-era assimilationist policies: this discourse of affirmative action comes not only from government officials on the ministerial level (ibid. 307), but also from the President of Tatarstan himself, Mintimer Shäymiyev. In a socialist-style propagandistic biography published in 1995, Shäymiyev’s biographers give a personal anecdote as one justification for his promotive language policies. In a section of the biography entitled *Tel – khalikning rukhi* ‘Language is the soul of a people’, Shäymiyev relates the story of his college entrance exams. At the time of the exams, Shäymiyev was bilingual but Tatar-dominant, and his entrance exams were given in Russian only. Apparently, for each question Shäymiyev would mentally translate the question from Russian in Tatar, think in Tatar, and translate his answer back into Russian, thus giving the appearance of a “slow-witted person” (Möstafin and Khäsänov 1995: 75). Shäymiyev passed his exams, but, regardless of the veracity of this anecdote, many other Tatars did not pass theirs: for years ethnic Tatars have been proportionally under-represented in Tatarstan’s institutes of higher education. Presumably it is incidents like these that inform the decisions and promotive policies of the Tatars who now rule Tatarstan. A few paragraphs after the anecdote above, Shäymiyev quite openly states his language policy priorities:

“Language,” repeats Mintimer Shäripovich, “is the soul of the people. The loss of your language means the loss of your national soul.¹⁵ We need to develop the Tatar language before anything else in order for there to be equal rights of languages. And this is not a violation of the principle of equal rights, only a restoration of historic justice. But it is necessary to work without rushing, thoroughly, and without compulsory measures. It is not acceptable to create conditions such as: learn the language in such and such amount of time, or we will replace you at your work. This might cause dissatisfaction, and discredit the idea itself. But it is necessary to create the proper conditions for speakers to learn the language; we need to open new schools and gymnasias where subjects are taught in Tatar, to introduce the mother tongue into the pre-school setting and into practice in the workplace, to publish textbooks in the mother tongue and to prepare specialists. This is not the work of one day: it is a long, complicated business that will take many years and decades” (ibid. 75-76).¹⁶

Some of the promotive policies of the Tatarstan government under Shäyminiyev (who has been president since 1990 and is currently in his third term) were discussed in section 1.1 above: the 1992 Law on the Languages of the Peoples of Tatarstan guarantees freedom of language use for all, bans discrimination on the basis of language, and promises that the republic’s government will protect the rights of Tatarstan’s citizens to develop their own languages (Graney 1999: 311). Two years later, a program with more than 100 policy recommendations was adopted by the Supreme Soviet of Tatarstan: almost all of these recommendations were aimed at the development of the Tatar language, and included calls for the increased use of Tatar as a medium of education and the establishment of a Tatar-language university (ibid. 311). This program also recommended ways for Russian monolingual government workers to learn Tatar such that government business could be transacted in either Russian or Tatar, including the publication of “minimum knowledge”

¹⁵ For more on Tatar discourse linking language, nationalism, and spiritual matters, see Chapter Two, particularly section 2.3.

¹⁶ This excerpt was taken from the Tatar-language version of Shäyminiyev’s biography. An identical Russian-language version was published simultaneously. All translations in this dissertation from Russian and Tatar, unless otherwise noted, are the author’s own.

Tatar dictionaries and basic Tatar classes for bureaucrats (ibid. 312). Work on these dictionaries must have begun immediately, as 1995 saw a spate of 50-page pocket Russian-Tatar work-related dictionaries with such titles as *Russko-tatarskii slovar' – Minimum dlia rabotnikov lesa* 'A Russian-Tatar Dictionary – The minimum for forest workers' and *Kratkii russko-tatarskii slovar' dlia rabotnikov ZAGSa* 'The short Russian-Tatar dictionary for ZAGs Workers.'¹⁷ Recall that in 1998, the Tatarstani government put into action a short-lived policy rewarding government workers for demonstrating a sufficient level of bilingualism: a contemporaneous report in a Moscow-based Russian newspaper was critical of this policy:

“Though this provision covers ethnic Russians as well, observers interpret it as a disguised discriminatory measure against them because virtually all the Tatars know Russian, whereas most Russians do not know Tatar. In reality, ethnic Tatars, who have declared themselves a title nation despite the fact that the Russian population of Tatarstan equals the Tatar population, will get a bonus just for being Tatars” (*Izvestia Press Digest* 1998).

A year and a half later, the English-language Moscow Times reported that Tatar-promoting affirmative action policies were not well-received by the republic's Russian population:

“Russians are beginning to grumble. They feel they are being gradually marginalized by policies that make it easier for Tatars to receive influential posts, as well as outright nepotism. Often, Russians are excluded from top posts because they don't speak the Tatar language, one of the republic's two official languages. ‘The channels of mobility are closed for Russians,’ says the Russian activist Salagayev, who is also a sociologist. ‘Russians are practically unrepresented in the legislative and executive organs.’ Indeed, a look at the names of the top figures in Tatarstan's government reveals a striking absence of Russians – who make up 43 percent of the population of Tatarstan and 54 percent of Kazan – among the elite. The overwhelming majority of Shaimiyev's administration officials, Cabinet ministers and State Council deputies are Tatar” (Karush 1999b).

However, the Tatar-speaking intelligentsia held, and holds to this day, the position that the Tatarstani government policies regarding language are *de jure*, but not *de facto*, that all government organs do not in fact conduct business and publish decisions in both Tatar and

¹⁷ ZAGs is the acronym of the Bureau of Records.

Russian, as promised by the 1992 legislation, and that the funds devoted to the development of the Tatar language have been to-date insufficient (Graney 1999: 311-2; also author's personal interviews). In sections 1.2.1 and 1.2.2, I will examine the effect that both Soviet-era and post-Soviet era language policies have had on the contraction of Tatar in both functional domains and population of speakers. An analysis of language attitudes and the functional domains where Tatar is and is not used gives evidence of continuing language shift: further motivation for language shift via accommodation is discussed in section 1.4, and further evidence of language attrition is given in Chapters 3 and 5.

1.2.1 Language attitudes

Many analyses of language shift have two commonalities: the first is what is called variously 'negative self-identification,' 'self-deprecation,' or 'low prestige,' and the second is resultant linguistic insecurity from this negative self-perception. Gal (1979), studying the shift from Hungarian to German in Oberwart, Austria found stigmatization and linguistic insecurity to be important motivating factors. She found a "near universal lack of linguistic pride and confidence on the part of the Oberwarters. Even those who can approximate both standards feel unsure and insecure about their speech when talking to monolinguals or outsiders" (ibid. 104). Gal asked informants to rank their language knowledge, with the local form being the standard: 50% ranked themselves below the highest rating for Hungarian, and 79% ranked themselves below the highest rating for German (ibid. 105). One Oberwarter stated, "We don't speak either Hungarian or German totally right" (ibid. 105). Dorian (1981) found a similar phenomenon among speakers of the highly stigmatized East Sutherland dialect of Scottish Gaelic (an already highly stigmatized language itself), who considered their dialect to be "deviant" and not worthy of study. Gal's Hungarian speakers were all associated with the

low-prestige agricultural peasant lifestyle, while Dorian's East Sutherland Gaelic speakers were descendants of poor tenant farmers who had been evicted en masse in 19th-century Scottish land reforms and placed in artificially created fishing communities, where they were stigmatized as well. The Tatar identity has also been one stigmatized for centuries: "Standard textbooks and historical literature have long contained negative images of the Tatar conquerors, in both Imperial and Soviet Russia; indeed, one pre-Revolutionary book, aimed at children and filled with anti-Tatar stereotypes, was republished several times in Russia in 1992 and 1993" (Shnirelman 1996: 55). Many Tatar speakers in present-day Tatarstan exhibit linguistic insecurity as well.

"...for urban Tatars the challenge is in not forgetting their traditional tongue. While today Tatar can be heard frequently on the streets of Kazan,¹⁸ during most of the Soviet period it carried a stigma. 'It makes me happy that children speak Tatar on the street. People of my generation feel insecure that we don't know the language well enough,' said Shavkat Altynbayev, who drives a gypsy cab in Kazan" (Karush 1999b).

None of the young Tatar speakers whom I interviewed gave themselves the highest rating for Tatar-language skills and many were hard-pressed to give me an example of a Tatar speaker whom they would rate a "10" for linguistic performance.¹⁹ Among the most common people cited were a prominent playwright and deputy in the Supreme Soviet and a prominent Tatar philologist; only one of these highly esteemed speakers, a member of the Tatar Social Club who works on the radio, was under 30 years old.

¹⁸ I take issue with the adverb 'frequently' in this sentence: Tatar is in fact heard *occasionally* on the streets of Kazan.

¹⁹ The exact wording of the questions was as follows: *Sinengchä, sin tatar telendä nindi därejädä söyleshä alasing? Äytik, ing äybät därejää un ball dip alsak, ä ing nachar därejää ber ball dip alsak, sineng därejäng nindi bulsa?* 'In your opinion, at what level do you speak Tatar? If the highest level is 10, and the lowest level is 1, what would your level be?' and *Sinengche, kemneng tatar tele un ball därejädä?* 'In your opinion, whose Tatar is on level 10?' The preceding questions had been about which language they felt most comfortable speaking, which language their primary, secondary, and college-level education had been in, and when they had learned Russian.

For many Russians, and also for many urban Tatars, modern Tatar is dismissed as “merely a kitchen language.” I somewhat informally tested language attitudes almost every day of my fieldwork in Tatarstan. The methodology of this informal testing was quite simple: I speak Russian with a foreign accent that, due to its extreme usefulness, I have not attempted to ameliorate. The number of tourists and foreigners in Kazan is proportionately quite low when compared with Saint Petersburg or Moscow, and merely speaking Russian with an accent (or, later in my fieldwork, speaking Tatar with an accent) was usually enough to cause my interlocutor to ask where I was from and what I was doing in Kazan. When people learned that I was there to learn and study Tatar, the universally asked question was simply “Why?”

In general, the Russians I spoke with were dismissive of Tatar as in any way meriting study, although there were notable exceptions. It was repeatedly asserted to me that Tatar was just a “kitchen” language, or the “language of peasants.” Some Russians seemed merely indifferent to the existence of the Tatar language, telling me things like “I’ve lived in this city 45 years, my whole life, and I don’t think I know even one word of Tatar.” Other Russians had been more attentive, and would comment on the frequent code-switching between Russian and Tatar that is a common style of linguistic performance for Kazan Tatars.²⁰ This behavior is often interpreted not as code-switching but rather as incomplete knowledge of or incompetence in Tatar – an interpretation that is most likely due to the monolingualism of the majority of Kazan’s Russians. One (monolingual) Russian said to me: “Those teenagers, they can’t say everything they want to say in Tatar, so they have to use Russian words in order to express themselves. They don’t know how to really speak Tatar.” Other Russians

²⁰ City Tatars of all ages can be heard both inter- and intra-sententially code-switching – I have observed people ranging in age from young school-children to elderly grandparents code-switching, most often with interlocutors who were their peers.

will openly express negative attitudes towards the Tatar language, particularly now that promotive language policies have brought mandatory Tatar language classes to primary and secondary schools. Several people, parents of school-age children, complained to me at length about the mandatory teaching of Tatar in schools as well as the teaching of Tatar literature that was apparently coming at the expense of the Russian canon. They deemed the language and literature classes to be “useless,” “a waste of time,” and “taking up time that should be used to teach ‘real’ things.”

Internal language attitudes, which is to say, those of Tatars themselves, varied widely among those people with whom I interacted, and there seems to be some sort of correlation with age and with Tatar proficiency.²¹ Older and proficient speakers usually reacted to my attempts to speak and ask about Tatar with delight. However, many city Tatars, particularly those under 30, seemed to regard the Tatar language with embarrassment, as is common for members of a minority community undergoing language shift. Once I had reached a reasonable level of competence in Tatar small-talk, I would attempt to use Tatar with all new acquaintances who had self-identified or been identified to me as Tatar. The response of many was of embarrassment, usually expressed by nervous laughter, a hand over the face, an inability to look me in the eye as I spoke to them in Tatar, and an unwillingness to speak Tatar in front of their peers.²² I received this response both from young Tatars whose close social networks were almost entirely Russian and from young Tatars whose close social networks were almost entirely Tatar.

By contrast, mostly under-30 attendees of the Tatar Social Club (described at length in section 1.3 below), almost all of whom are actively engaged in Tatar culture and/or

²¹ A more precise correlation must wait for future fieldwork.

²² I can only guess where they are located on the cline of language competence, since they would not speak with me in Tatar.

politics, reacted incredibly positively, almost overwhelmingly so, to my attempts to learn and speak Tatar. Several, upon acquaintance, immediately took it upon themselves to offer to meet with me regularly and help me with my studies so that I would learn to speak well. This was true of the Tatar intellectual and cultural elite in general, among whom Tatar has “covert prestige” (Trudgill 1972) and is a deeply emotional topic. For example, two young women whom I met in the club (approximately 20 years old at the time) informed me on separate occasions that hearing me speak Tatar made them want to cry, although they weren’t sure why. I have heard from fieldworkers studying other low-prestige or endangered languages about their experiences where speaking just a few words of the minority language caused tears to well up in the eyes of their interlocutors – but the interlocutors in question were always grandmothers, and never college students.

1.2.2 Tatar’s functional domains

While the Tatar and Russian languages have in principle been on equal footing since 1992, there is a distinct asymmetry in their functional domains and usage, particularly in urban Tatarstan. While Tatar usage has greatly increased since the declaration of sovereignty, Russian is still, by far, the dominant language, and the language of the public domain. The vast majority of newspapers, magazines, television shows, and radio programs are in Russian only, as are government proceedings and the majority of educational opportunities. Discontent with this asymmetry, and with the concession of both public space and public discourse to the Russian language, is found in articles in the post-Soviet Tatar press with titles such as *Tatar tele chinnan da däülät teleme?! ‘Is the Tatar language in reality an official language?’* (Fättakh 2000), *Tatarcha sorıybız, ruscha jawap alabız ‘We ask in Tatar, and receive an answer in Russian’* (Maksimova 2000), and *Tatar ilendäge kiosklarda nigä tatar gazetalari yuk?*

‘Why are there no Tatar newspapers in the kiosks of a Tatar country?’ (Mirsäyetov 1999). And one author, who describes her decision to study in Tatar at the university level as *batırlık* ‘heroism’, states: “In 1992 the law ‘About the Languages of the Peoples of Tatarstan’ was passed. Thus, Tatar and Russian both became government languages, and these two languages were given equal rights. But...when two Tatars are standing and conversing and just one Russian person arrives, why does the conversation switch to Russian?”(Khäyrullina 1999).

Tatar accommodation to Russian is due, in great part, to Russian monolingualism: there is a distinct asymmetry between Russian use and usefulness and Tatar use and usefulness. As noted above, the Soviet Union was famous for its asymmetry of bilingualism, where minorities learned the dominant language and Russians did not; Tatarstan is no exception. A 1989 survey showed that 1.1% of the republic’s Russians spoke Tatar, while 77% of Tatars spoke Russian (Walker 1996).²³ Statistics on Tatar knowledge are less credible: statistics are bandied about that show that 99% of rural Tatars and 95% of urban Tatars claim that Tatar is their native language (cf. e.g. Iskhakov and Musina 1991). However, it is important to note that neither Russian *rodnoi iazyk* nor Tatar *tugan tel*, both of which translate to ‘mother tongue’ or ‘native language’, require either knowledge or use of the actual language: they merely refer to the language of heritage traditionally associated with the ethnicity. It is possible to say in both Russian and Tatar, “She does not know her native language”: therefore, statistics on what Tatars consider to be their “native language” should

²³ These statistics imply that 23% of the republic’s Tatars were monolingual in Tatar in 1989. Based on (admittedly) anecdotal evidence gathered just eleven years later, I find that number to be rather high. I personally never encountered any monolingual Tatars, even in villages, and never heard of any monolingual Tatars under the age of 80.

not be read as measures of linguistic performance, but rather as symbolic statements of ethnicity. In fact, a 1989 survey stated that only 36% of urban Tatars in Tatarstan used Tatar as a home language, 25% spoke Tatar when socializing with friends, and 21% spoke Tatar when interacting with colleagues (Graney 1999: 310).

In Kazan, the main functional domains for Tatar are as follows. Tatar is the home language for some, particularly for those whose parents moved to Kazan from a village as part of the upsurge in urbanization in the 1970s. Tatar can also be the language of education. Recall that by the end of the Soviet period only 12% of Tatar children were in schools where Tatar was the medium of education, and the vast majority of these schools were not in urban regions. For years, the Tatar department of Kazan State University was not-so-affectionately nicknamed “the *kolkhoz*” ‘collective farm’ due to the overwhelmingly rural origins of its students. Prior to 1990, the proportion of rural to urban students in the department was approximately 95 percent to 5 percent.²⁴ Today the department’s students are approximately half rural and half urban, and it is claimed by the dean of the Tatar *fakul’tet* that this change is due to the increase in the availability of Tatar-medium education in the cities (Iskändär Gilyazov, personal communication). In Kazan, parents often have a choice of school, and can send their children to schools where Tatar is the medium of education (often in conjunction with others, in particular Turkish and English) or where Russian is the medium of education, although according to Lotfullin (2000), in Kazan there are currently five times

²⁴ Tatar “retreated” to the villages in the 20th century, which were considered Tatarphone strongholds (hence the association between Tatar and “peasants”). However, Russian has begun encroaching in the domestic domain even in villages, particularly for school-age and college-age Tatars, who will code-switch heavily in social situations, even if Tatar has been both the language of home and of education.

more Russian-language schools and teachers than Tatar.²⁵ In Kazan at the university level, there are currently five places where one can study in Tatar: (1) Islamic University (and Tatar as non-native language is taught there, with students at all levels of proficiency); (2) The Tatar department of the Pedagogical Institute; (3) The Tatar department of Kazan State University; (4) One group (out of four) each year in the Journalism department of Kazan State University; and (5) several departments in the Tatar Humanities Institute, which was founded quite recently and whose student population is thus rather small. The majority of students in Kazan who are studying in Tatar are being trained to be teachers (for primary, secondary, and university-level education), philologists, and journalists. All other education at the university or institute level is conducted in Russian. In addition, students in one of the above departments will take “general” courses, such as sociology, history of Russia, or economics, in Russian.

Tatar, while present in the media, is once again asymmetrically represented, where the vast majority of news and entertainment is available in Russian. At the time of my fieldwork there were four radio stations that broadcast in Tatar: (1) Radio Dulkan, the “traditional” Tatar station, which broadcast in Tatar only,²⁶ (2) Radio Kurai, which was aimed at a younger audience, playing Russian music occasionally and allowing DJs to periodically code-switch into Russian, (3) Tatar Radiosi, which began broadcasting in August of 2001 and plays music in Tatar, Russian and English, with DJs speaking set phrases in all three languages, and (4) Bolgar FM, a government station, which broadcasts twelve hours of Tatar a day in three-hour slots – the other twelve hours are Russian-language programming. Additionally, another government station broadcasts several hours a day in Tatar, and

²⁵ In rural areas there may be no choice of school, and thus no choice of language of education.

²⁶ Radio Dulkan lost its license in the spring of 2002 and is no longer broadcasting. Its replacement apparently will broadcast in both Tatar and Russian (RFE/RL, April 1, 2002).

includes *Azatlík* 'Freedom', the Tatar-language broadcast of Radio Free Europe. This station cannot be found on regular radios.²⁷ All of the remaining stations, approximately eight, broadcast in Russian only and do not play Tatar music, although they will play American, European, and Turkish music.

On television, there is one government-run Tatar station which broadcasts only six hours a day, and of these six hours typically only three are in Tatar. The other local station, *Efir*, now offers three Tatar-language programs each week, two of which began in 2001. This asymmetry is continuous with Soviet-era asymmetries in media production, where Russian-language television programming was available at all times, while local broadcasts, radio or television, were "of a highly restrictive kind" and distinctly amateurish, due in part to their comparatively low budgets (Ozolins 1996: 186); the same is true today.

The majority of newspapers and magazines found in Kazan are in Russian. Some newspapers will put out both Russian and Tatar versions, but it should be noted that these different versions will have completely different editorial staffs and almost completely different contents. Reading *Tatar Ile* 'Tatar Country' in Tatar is not the same as reading *Tatarskaya Kraya* in Russian. Not only are there fewer newspapers printed in Tatar than in Russian, in part, because press from outside the Republic is in Russian only, these papers are in general quite difficult to find. Most neighborhoods in Kazan have a series of outdoor wooden bulletin boards where newspapers are posted, so that those people who can't afford to buy them can read them away. The newspapers are usually grouped in threes or fours, and the proportion of Russian to Tatar papers is usually 3 to 0, 4 to 0, or 3 to 1. In addition, it

²⁷ Every apartment and house is furnished with the kitchen "government radio," a one-channel radio that plays a government station (until *perestroika*, the only option for radio).

can be difficult to find a place to buy a Tatar newspaper: they are not available at the newsstands at the Kazan train station or bus station, and the prominent newsstands and kiosks downtown also do not carry them.²⁸ Tatarphones know via word of mouth where the few downtown kiosks that sell Tatar newspapers are located, but more usually, they subscribe to the papers and have them delivered. The same is true of Tatar magazines: they must be purchased by subscription, and newsstands that sell magazines will have only Russian-language publications.

Finally, a note on quality. While Tatar language magazines²⁹ appear to be on a par with Russian-language magazines, in general, Tatar television programs and Tatar newspapers are of significantly lower quality than Russian television and newspapers. I know many people who actively participate in Tatar culture (philologists, theater workers, club members) who refuse to read the Tatar press (or to teach in the Journalism Department of Kazan State University) because they feel that the standards are so low, and who instead get their news from a Russian-language newspaper (although perhaps one written mostly by and for Tatars, such as *Vostochnyi Ekspres* ‘The Oriental Express’). Others will not watch Tatar-language television programs, with the exception of the news on Tatar Republic Television (TRT), which is the exact Tatar-language equivalent of its Russian counterpart. I have seen people turn away from a television playing a Tatar show in embarrassment, saying “please turn that off, I can’t stand to watch it.”

The asymmetry of Tatar and Russian proportions is even more pronounced in daily public activities in the city. The public domain appears to be a Russian-only one, and Tatars

²⁸ This is described in detail in Mirsäyetov 1999.

²⁹ There are four major magazines, three of which are continuations of Soviet-era magazines: *Kazan Utlari* ‘The Fires of Kazan’, a literary journal; *Söyembike*, a woman’s magazine; and *Yalkin* ‘Flame’ a children’s/teen’s magazine. The newest Tatar magazine is *Idel* ‘Volga’, a youth-oriented literary journal.

will often accommodate and speak Russian when in public spaces or engaged in activities in the public realm. Although Tatar is an “official” language, many Tatars in both interviews and conversations expressed an opinion that this officialness is *de jure* only, and there is a sense that people are only “playing Tatar,” which is to say, presenting only enough Tatar to give an appearance of compliance and language equality (cf. e.g. Makhsimova 2000 and Fättakh 1998). For example, there is legislation requiring equal public signage, but this legislation is not enforced.³⁰ The main signage for businesses and government buildings is usually bilingual, with precedence given to neither language: the description of the store’s business and the office hours are always found in both Russian and Tatar. However, with the exception of places under government jurisdiction such as the post-office or library, or government-run stores such as *TsUM*, the main department store, and bread stores, all other signage within a public establishment will usually be in Russian. Price tags, sale signs, descriptions of merchandise, elaborations on working hours, policy signs: these are all in Russian only. In addition, purchases are made in Russian only. The same is true on transportation: the main bus and tram signs, describing the route, the price of a ticket, and the fine for avoiding a ticket are in both Tatar and Russian, but all other signage is in Russian. Transactions on public transportation are also in Russian only, and on several occasions I saw or was told about Tatars attempting to buy tickets in Tatar who were scolded by Russophone conductors or tram drivers. In addition, I observed the following situation multiple times: two salespeople are speaking Tatar behind the counter in a store,

³⁰The Russian-language television show *Gorod* ‘City’ on the local channel *Efir* did a not-terribly-serious 10-minute exposé of this lack of compliance in April of 2001. The reporter went from store to store showing Russian-only signage and price tags, and asked mostly indifferent clerks for explanations and justifications, to no great effect.

two Tatars walk up to the counter, also speaking Tatar. The conversations pause, the official transaction is made in Russian, and then both Tatar-language conversations resume. The one major exception to this rule of accommodation to Russian in public space is the downtown produce market, the *Kolkhoznyi Rynok* ‘Collective Farm Market’, which was one of the few public spaces where I consistently heard Tatar spoken. I received the following explanation for this phenomenon: this market is the cheapest market in Tatarstan with the widest selection of goods, and Tatar-dominant villagers, people who prefer to speak Tatar and have a higher level of Tatar competence than Russian, travel into the city in order to shop there. The economic force they represent is sufficiently powerful to make learning market-Tatar desirable for workers there: not only do the Uzbek, Tajik and Azeri sellers transact business in both Russian and Tatar with their customers, but also it is possible to find Russians who are making efforts to speak and understand some Tatar.

However, the professions where one is able to use Tatar as the medium of work are rather limited, and disproportionately tilted towards cultural and media production: theater, music, television, radio, magazines, and education, both secular and religious. Salaries for these professions are, in general, quite low: higher-than-average salaries are not found in the arts or education, but in business, particularly in businesses with Western connections, and in these jobs Tatar knowledge is generally not an advantage.

Finally, despite promotive language and cultural policies on the part of the Tatarstani government, an examination of public discourse demonstrates that there is a sense that Tatarphones feel themselves to be “under siege” from Russian cultural and linguistic pressures. In the Tatar press one finds almost weekly articles and letters to the editor expressing this siege mentality, for example: *Shovinstlar berläshä, ä bez haman yokläybiz* ‘Chauvinists are uniting, and we are permanently sleeping’ (Äkhmätjanov 1994), *Bez, tatarlar,*

yäshibeꞗ äle! ‘We Tatars, we live after all!’ (Galimullin 1995) *Telebezne, dinebezne sakliyk!* ‘Let’s save our language, our religion!’ (Mäjitoꞗ 1996). There is a derogatory name for Tatars who have assimilated to Russian culture: “Mankurts”, a name taken from a novel by Chingiz Aitmatov. I first heard the term in a bar, when someone in the group I was with spotted an acquaintance of hers sitting with his Russian girlfriend and friends. “Mankurt!” she muttered, and when asked for an explanation, she said, “Mankurts, they are Tatars who don’t know their own language or history, their own traditions. They don’t understand or respect their ancestors.” In keeping with the metaphor of the Tatar nation and culture under siege, Mankurts are perceived and referred to as “traitors” who have “defected to the other side.”

So where can one place Tatar in the process of language contraction? Fishman (1991: 81-121) presents a graded typology of language contraction situations that is designed to be used for the process of reversing language shift. He labels his typology the Graded Intergenerational Disruption Scale (GIDS), and “the higher the GIDS rating the lower the intergenerational continuity and maintenance prospects of a language network or community” (ibid. 87). This scale is a “quasi-implicational scale,” such that higher levels of the scale imply the inclusion of all the disruption types found in the lower levels of the scale.

There are eight stages in the typology of disruption, which are presented in descending order of level of contraction.

- Stage 8. An advanced degree of language attrition, with the only speakers being socially isolated members of the oldest generation who usually do not have conversational partners in the minority language; a language in this stage would be deemed “moribund” (Krauss 1992). Fishman includes speakers with all levels of competence, including those that Dorian (1981) would classify as “semi-speakers” and as “rememberers,” and the oral culture preserved by them may be merely “salvageable.”

- Stage 7. Speakers of the minority language are beyond child-bearing age, but are socially integrated in an ethnolinguistically active population; this integration is the main difference between Stage 7 and the linguistic isolation of Stage 8. In this stage of language contraction, the older minority-language speakers can be mobilized to change the linguistic behavioral patterns of children, and thus work around the disruption of normal language transmission in the domestic domain.
- Stage 6. The minority language is transmitted intergenerationally, with daily informal oral interaction between the generations, and has a modicum of institutional reinforcement in addition to demographic concentration. The “lion’s share of the world’s intergenerationally continuous languages are at this very stage and they continue to survive and, in most cases, even to thrive, without going on to subsequent (‘higher’) stages” (ibid. 92.). This stage is the crucial stage of language maintenance and language revitalization, and can not be skipped in the process of revitalization. Stage 6 represents stable diglossia, with clearly assigned functional domains for the minority language (almost completely domestic) and the dominant language (all other domains), and all the stages below 6 also represent stable diglossia, simply with expanded minority-language domains.
- Stage 5. Intergenerational oral communication in the minority language is supplemented by minority-language literacy in the minority community.
- Stage 4. In addition to intergenerational oral communication in the home and literacy in the community, there is minority-language education at the lower levels of the educational system.
- Stage 3. In addition to the functional domains of Stage 4, the minority language is used in the “lower work sphere,” even in places outside the minority community.

- Stage 2. In addition to the functional domains of Stage 3, the minority language is available in lower governmental services and lower spheres of the mass media.
- Stage 1. In addition to the functional domains of Stage 2, the minority language is used in higher-level education, in the general media, and a range of workplaces, and in a variety of governmental spheres, although the minority community does not have political independence.

It is difficult to classify Tatar in Fishman's graded scale, as it is simultaneously contracting and expanding. Most importantly, there appears to be a disruption in intergenerational transmission as described in Stage 6: the domestic sphere of Tatar usage is not a stable one, and the language of choice for intragenerational interaction for the youngest generation is often code-switched Russian and Tatar, or entirely Russian speech. Therefore, Tatar is not in a situation of stable diglossia, but is rather a contracting language. Even so, there are elements of the earlier, more stable stages of minority-language usage to be found in the Tatar case, both because pre-Revolutionary Tatar was a strong language with influence beyond the Volga Tatar community and vestiges of its pre-Soviet functional domains remain, and because promotive language policies (and the cessation of anti-Tatar Soviet policies) have caused an expansion in Tatar's functional domains. Therefore we see the minority-language literacy of Stage 5, the lower-level minority-language education of Stage 4, elements of minority-language in the workplace, found in Stage 3, *de jure* availability of Tatar in lower-level governmental spheres and real availability in the media, found in Stage 2, and higher-level education in Tatar, found in Stage 1. However, without the continuation of intergenerational transmission, the foundation for the other functional domains of Tatar found in the other stages is shaky at best; if the population of Tatar speakers continues to contract, the functional domains and uses of Tatar will eventually contract as well.

1.3 The “Tatar Social Club”

My focus in this dissertation is on the linguistic performance of young (mostly college-age) bilingual Tatars living in Kazan. I made the acquaintance of these young Tatars through a club that met either weekly or bi-weekly in downtown Kazan, a club that I am calling the “Tatar Social Club.”³¹

The Tatar Social Club was founded in 1997 by three culturally and politically active Tatars in their early and mid-twenties: two of them were studying in the same department and the other was prominent in Tatar political circles. At first, meetings were small and haphazard, but the club began to meet more regularly in 1998, and in the 2000-2001 season (which runs parallel to the academic year) the Tatar Social Club met twice a month in the fall and nearly every week in the spring. While in principle “every attendee is both host and guest” at a club meeting (as stated in the club’s promotional materials), in reality there are several core organizers who keep the club running. Two of the three original organizers have retained key organizational roles, and three more organizers joined their ranks soon after the club’s founding: these organizers are all part of a “dense” social network (Milroy 1987) and have ties through a variety of Tatar political, cultural, and religious organizations.

Club meetings, which take place on weekends, are attended by between 20 and 100 people, and in the spring of 2001, there were an average of 50-60 per meeting. There is a wide range in age of people attending meetings: some people bring their young or teenaged children, other young Tatars are accompanied by a culturally involved parent, and a few

³¹ As an added level of protection for the young Tatars who allowed me to work with them, I am using a pseudonym for the club. Therefore I will not give references for the articles about the club that are cited in this section.

older members of the Tatar intelligentsia will sometimes stop by. However, the range in age of the core club members in 2000-2001 was between 18 and 30. I use the term “club member” for ease of reference only: there are no membership requirements or membership lists and people are neither invited to join the club nor expelled from it.³² However, there is talk of what is “our type” and “not our type”, and while anyone who hears about the club and is interested is welcome to attend, the core club members are a self-selecting group with shared characteristics. These core club members are also part of a dense social network and have multiple ties outside of the club: they will perhaps live in the same dormitory (some dormitories are residences for students and graduate students of more than one institution), study in the same *fakul'tet* (a organizational level above that of ‘department’), or have connections through their work (at the same newspaper, or by means of working for organizations that communicate with one another). Some young Tatars come to the club without any personal connections, having heard about it on the radio or read about it in the newspaper, but these people without additional personal ties rarely become core club members. All of the club organizers are of rural origin, as are the majority of core club members: they have come to Kazan for their education, and those who have graduated have stayed in Kazan for work opportunities. One club organizer periodically considers changing the day of meetings to a weekday, as he feels that the current weekend meeting time excludes many potential club members, students who live in the city dormitories during the week and commute home to their villages for the weekend (this is the reason that the Tatar disco held

³² One club organizer told me that he was a little regretful that people could not be asked to leave, as there was an occasional female attendee whom he characterized as of questionable morals and behavior, who “associated too freely with men” and was “impudent”; he wanted to ask her to stop attending meetings, but wasn’t able to, due to club policy.

at *NKTs*, a national museum located in downtown Kazan, is held on Thursdays rather than on Saturdays).

Attendees of the Tatar Social Club are not members of Tatarstan's socio-economic elite, but they do tend to be members of the Tatar cultural elite, and in articles or radio pieces about the club, they are often referred to as the "cream of Tatar youth." Those who have graduated from college often work in high-prestige but low-paying professions: teachers, philologists, and journalists in particular are over-represented in the club. While quite young, club attendees work in the highest level of the Tatar cultural world, and hold important positions in Tatar theatrical circles, at Tatar newspapers and magazines (both as writers and editors), and as on-air personalities on both Tatar television and radio. Club members tend to not be well-traveled – many have not been outside the borders of Tatarstan, let alone the Russian Federation – and they often have a more "provincial" or Tatarstan-oriented world view. One club organizer in a published interview in the Tatar press characterized club attendees in this way:

"The youth in our club hold traditional views. This is understandable; 'westernized' youth don't come to us, they prefer striptease bars, nightclubs, alcohol, beer, narcotics, etc. Logically, we should have become a striptease bar...but it seems that there remains among the Tatar people an instinct for self-preservation. Quite a few of these young Tatars engage in creative activities, and they in some measure prefer Islamic values to a western lifestyle."

This statement is notable both for its caricature of Kazan's 'westernized' youth and its assertion that club members are religiously oriented and morally conservative. It also makes very clear who is welcome at the club and who is not: overtly sexually active, religiously unobservant, western-oriented Tatar youth are presented as "not our type."

While this characterization of club members as (by implicit contrast) chaste, traditional, non-Westernized, religiously observant teetotalers is exaggerated, these are the values that are publicly presented as ideals in the context of club meetings. The club is an

important locus for the construction of Tatar identity, particularly in its idealized form, and values that are seen as incongruous with this idealized Tatar identity are not expressed in the forum of club meetings or club outings. Alcohol is not served at the club, and those club members who do drink alcohol do not openly discuss it at club meetings. Sexual activity and alternative sexuality too are topics that are not openly discussed: for example, one club organizer would regularly “protect” the reputation of a former club member by referring to her former apartment-mate as a female friend rather than her boyfriend,³³ and gay and bisexual club attendees who go to Kazan’s gay bar hide this fact from other club members.

However, the focus of the club is not really on personal mores: club members are linked by their cultural and political activism. The club is financially independent (there are essentially no expenditures), and is unaffiliated with any political or religious organizations: club members have refused offers of financial support both from the Tatarstani government and from oppositional organizations. (Other Tatar cultural venues for students such as the *Yagimli Yaz* ‘Sweet Spring’ festival have government support that runs into thousands of rubles.) However, the club’s unofficial political orientation is quite clear: many of the most prominent club members are nationalist separatists, although they may belong to different nationalist movements (pro-Shäimiyev, anti-Shäimiyev, pan-Turkic, etc.). Club members are highly visible presences at political rallies and protests, and can be found en masse at the *Khäter Köne* ‘Memory Day,’ which commemorates the fall of the Kazan Khanate to Ivan the Terrible in the fall of 1552. The club organizes a yearly “pilgrimage” to Bolgar, a town south of the confluence of the Volga and Kama rivers that was the seat of the Volga Bulgaria, a

³³ She herself was quite open that she had been living with her boyfriend, and cited one of the reasons that she had stopped attending the club as its “hypocrisy.”

political predecessor to present-day Tatarstan (discussed in section 1.1 above). Although the 10th century ruins found at the site are generally of a secular nature, it is considered a “holy” place, and out of respect, female club members don headscarves. One club member brings the three flags that symbolize his political orientation with him on this trip: a Tatarstani flag, a Turkish flag, and the flag of the Idel-Ural movement, a pale wolf on a green background. These three flags, which are also prevalent at the *Khäter Köne* ceremonies, are waved out the bus windows en route to and from Bolgar, and club members often pose for group pictures where the flags are prominently displayed. Finally, the meetings themselves take place on the second floor of a building that houses a Tatar nationalist organization: the banner on top of this building, sitting on the roof just feet above the heads of club members and facing one of the main thoroughfares of the city, states: *Makhsatibiz: Tatarstanning Bäjsezlege* ‘Our Goal: Tatarstan’s Independence.’

The Tatar Social Club is considered by its organizers to be the continuation of an early 20th century (pre-Revolutionary) club of the same name. This pre-Revolutionary club was in fact more like a cultural salon, and among its small circle of members were Tatar intellectuals who were the main producers of Tatar culture of the time, particularly literature, music, and theater. By choosing to use the same name, the club organizers imply that after a hiatus of 90 years, they are returning to the legacy of the intellectuals of the “golden era” of Tatar cultural production (generally considered to be from the mid-19th century to the Russian Revolution). However, when asked why the club was founded or what its purpose is, club organizers and members do not cite a devotion to Tatar culture and cultural production. Instead, the first (and sometimes only) answer given is a linguistic one: the club was founded as a place where Tatar youth could socialize in Tatar, and can thus be interpreted as an attempt to carve out a new Tatar domain in the overwhelmingly Russian-speaking context of

urban Kazan. Written discourse on the club, both in the club's promotional materials and in articles about the club, also explicitly links the club to being Tatarphone. One club organizer, asked in an interview about the club's purpose, called it a place where "Tatar-speaking youth gather." Another article about the club begins this way:

"Who are you, reader? Are you a person who has been wanting to learn Russian your whole life, who has started to forget your Tatarness? Are you a child who has left behind the words of your favorite grandmother? Do you study in one of the capital's Tatar schools, gymnasia, specialized high schools, technical schools, institutes or universities? Whoever you are, and however you are, you are our person: a Tatar child."

Once again we are given characteristics of people who are welcome at the club, "our" type: here it is an interest in not forgetting either one's Tatarness or language of heritage. Yet another article about the club, this one in a Russian-language Tatar newspaper, claims that the club is popular "because it is a place in Kazan where it is possible to socialize in one's native language, feel oneself equal among equals." Here we have allusion to both the particularization and stigmatization of Tatar. The club is characterized as an in-group setting free of stigmatization, and a significant part of being able to feel "equal among equals" is the club's explicitly Tatarphone setting. New promotional materials for the club, written after the period of my fieldwork, state: "No matter what the theme, meetings begin and end with the song 'I Tugan Tel' [O Native Language] which symbolically expresses our sincere love for our native language."³⁴ This practice of beginning and ending each meeting with the unofficial Tatar national anthem began a few months after my fieldwork ended in the summer of 2001, and I observed it on a follow-up visit in the spring of 2002.

Club meetings had a predictable pattern. Each meeting would have a different host (either a club organizer or a person assigned to host the evening) who would have arranged the evening's activities and invited guest speakers or performers, if there were any. Meetings

would have a theme chosen in advance by the club organizers: themes for the 2000-2001 season included evenings devoted to Tatar populations outside of Tatarstan (both Siberian Tatars and the Tatars of Orenburg), the “friendship” between Tatars and other ethnicities (there were Jewish, Arabic, and Turkish “friendship” evenings), intellectual topics (Conversativism vs. Jadidism, the role of Islam in the modern world, the work and influence of the poet Tukay), and youth-oriented topics (how to travel by hitchhiking, how to use the internet and what Tatar resources can be found there). In addition, there were several musical evenings, including a concert for the spring New Year’s holiday, *Näürüz*, which was broadcast live on Tatar radio, and an amateur concert put on by club members themselves. Before each meeting began, starting at around 6:30, men would arrange the furniture to suit the number of people who had come, and women would begin to prepare tea and plates of snacks (attendees were asked to bring store-bought sweets). A club organizer would introduce the host, who would then act as emcee for the rest of the evening. They would introduce the evening’s topic and guest speakers or performers, and then the performance would begin. After a certain amount of time, determined both by the host and by the long-windedness of the guest speakers, the host would call an end to the first part of the evening. A club organizer would then take over and begin the process of self-introduction: each person attending the meeting would stand up and introduce herself or himself with some basic information: their name, their place of origin, their year and department of study or profession and workplace. After the last introduction, announcements would be read, and then a break would be called. After the break, which would last about 45 minutes, the evening would continue until about 9:00 p.m., usually with a brief continuation of the pre-break theme and then renewed socializing.

³⁴ The poem and song *I Tugan Tel* are discussed at length in Section 2.3.1 of Chapter Two.

All club-related speech would be in Tatar, both the public speaking of the hosts and guests, and the private speech of attendees engaged in activities like tea-making and furniture-moving. While outside the club Tatar speakers will almost always accommodate people who express a preference for speaking in Russian by using Russian with them, during club meetings there are no concessions made. The club is a Tatarphone space, and use of Russian, even by speakers who are not fully competent in Tatar, is both implicitly and explicitly discouraged. This can be traumatic for some during the self-introduction portion of the evening, and less-competent Tatar speakers will cope with different strategies: some will have a fluent friend who will apologize for their lack of Tatar competence and do the introduction for them, others will flee downstairs as the introductions begin and wait out the process, and over the course of the year I saw a few attendees introduce themselves in Russian and begin crying as they explained that they were unable to do so in Tatar. During club meetings, core club members will refuse to accommodate locals and foreigners both, even when Russian is available as a medium of communication. I attended my first club meeting just three weeks after my arrival in Tatarstan, when I had only the most limited Tatar competence, but the club organizers who immediately befriended me would not speak to me in Russian within the physical and temporal space of the club. They might speak to me in Russian during the break period, or if we had moved downstairs or outside, but during the boundaries of club meeting time and space, ideology was more important than communication and they would speak Tatar only. This refusal to accommodate could perhaps be interpreted as a desire to help me learn Tatar faster; however, foreign guests of club members, people who had expressed no interest in learning the Tatar language, were also addressed exclusively in Tatar during club meetings, while the club members who had brought them would translate for them. Urla (1987: 301) notes a similar lack of

accommodation to outsiders among Basque speakers; for example, when her sister, a fluent Spanish speaker, visited her at the fieldwork site, the local Basques would address her in Basque only and wait for Urla to translate both their speech and her sister's response. Once I became somewhat more fluent in Tatar, club members would treat me the same way: I started bringing a young American studying in Kazan to club meetings in order to introduce him to people I thought he would like, and during meetings, although he is perfectly fluent in Russian, he was addressed in Tatar only, and I was expected to translate both questions addressed to him and his responses. Some club members would continue this behavior outside of the club – one club organizer in particular would never use Russian with my American friend, even though he had quite freely spoken Russian with me when we were becoming acquainted, both in order to facilitate my Tatar-language learning and because at that moment communication had been a priority for him, both to determine my intentions and character, and, once they had been approved, to aid me in my goals.

It has been noted that people are reluctant to change the language of their acquaintanceship (cf. Urla 1987: 304ff): this is one justification for full immersion classrooms in second-language teaching, to create the habit of speaking that second language. However, the members of the Tatar Social Club were not only willing but eager to change the language of their association with me to Tatar as soon as possible (sometimes even sooner). Some would explicitly announce the change in the language of our acquaintanceship. A month into my fieldwork, one woman took me to her aunt's village in the north for a weekend (her own village was too far away) and announced en route: "From this moment on, we're not going to speak Russian, we're going to speak Tatar." She made the same announcement to her family upon arrival, and they proceeded to speak to me in Tatar only for the entirety of my stay, even though I was generally uncomprehending; not only were there dialectal differences

from the Tatar that I had been studying, but they also spoke at what seemed to be twice the speed of urban Tatars and didn't have any kind of "foreigner talk" with which they could accommodate me (as the occasion for it had probably never arisen previously). My friend would translate from their Tatar into a simplified Tatar, and did the same when we hitchhiked back to Kazan with three older village Tatars who, although bursting with curiosity to speak with the first American they had ever encountered, accepted the implicit rule that our conversation was to be in Tatar only, and patiently waited while my friend played her role as Tatar-to-Tatar translator. This woman took it upon herself to introduce me to various aspects of Tatar culture: she took me to concerts and to plays (where she wouldn't let me use the simultaneous translation), invited me to the office of the Tatar-language publication where she worked, and taught me how to cook traditional Tatar dishes – and she never spoke to me in Russian again.

I encountered this Tatar-to-Tatar translation phenomenon several times more during my fieldwork, particularly when people would invite me home to their village where the dialect was unfamiliar. Their implicit reluctance to speak to me in Russian, or to allow me to speak in Russian, was made explicit by some other young Tatars of my acquaintance. I had regular Tatar-English conversational exchanges with a few club members, where we would converse for an hour in Tatar, an hour in English, and then socialize afterwards – this socializing would just be another version of the Tatar conversation hour, because I was not allowed recourse into Russian. If I would stumble in Tatar and revert to Russian to express myself, my conversation partner would request, "Say it in English or Tatar, please," even if the original language of our acquaintance had been Russian. Some other more moderate members of the Tatar Social Club simply stopped speaking to me in Russian, presumably after deciding that my Tatar was up to snuff. There were two club members in particular

whom I did not see for a space of several months: the language of our association before the hiatus was Russian, and afterwards was Tatar only. When this was brought to their attention later, they both claimed to not really remember speaking with me in Russian: “Didn’t we always speak in Tatar?” one asked. By the end of my stay in Tatarstan, all of the Tatars who considered themselves to be competent speakers used Tatar to converse with me, regardless of the original language of our acquaintanceship, and I almost never initiated the change in language.

During the course of my fieldwork, I observed and got to know approximately 40 club members somewhere between reasonably and very well, and was acquainted with approximately 60 more young bilingual Tatars. I engaged in standard participant-observation activities: I participated in daily life, hung out in dormitories, danced in the Tatar disco, attended religious festivals in homes, went to mosques, was invited on village excursions, visited other cities, and once I had moved into my own apartment, hosted many impromptu social gatherings. All of my observations were recorded in field notes.

I used two methods to collect information on language attitudes and other sociolinguistic data. The first was informal elicitation during seemingly casual speech (which would then be recorded in field notes). The second was recorded sociolinguistic interviews composed of 52 questions, eliciting information on self-reported language use, language awareness, opinions on the new Latinized alphabet, Tatar-language education, and more.

I gathered conversational data in a variety of ways. The first was through conversations taped in my kitchen, where people who are usually conversational partners agreed to come to my house and speak for my tape recorder, after which we socialized. In an attempt to counter the “observer’s paradox” and element of artificiality found in these recorded conversations, I began to ask permission to record people who were already

gathered in my kitchen, people who were there “just to hang out.” This was only possible after my Tatar had become good enough to participate in the conversations without excessive accommodation on the part of the conversation participants. My final recording method was the “home recording kit” – after searching Kazan exhaustively for dictaphones I put together kits that included a small tape recorder, two cassettes, permission slips,³⁵ and instructions. I then asked my informants to tape conversations that were taking place in their usual place and usual time: at home around the dinner table, in the dorm having tea with friends, etc. However, despite my efforts, all of my recordings turned out to be in an “on-stage” style: in the next section I will analyze the role of the fieldworker in style shifting and data collection, and show how this contextualized data can be used in both sociolinguistic and grammatical analysis.

1.4 Fieldwork, the “observer’s paradox,” and data contextualization

In this section, I will examine some of the difficulties faced by the linguistic fieldworker who is attempting to observe and record “natural” conversations, and I will reconsider the long-held sociolinguistic notion of the observer’s paradox by recasting it within Bell 1984’s framework of audience design theory. Using data gathered during my own fieldwork, I will once again call into question the idea of a single, unmarked, unperformed vernacular, the access to which is supposedly blocked by the observer’s paradox. Finally, I will demonstrate that “performed” or “self-conscious” speech produced for the fieldworker can be useful in

³⁵ In accordance with the protocol established by the Committee for the Protection of Human Subjects at the University of California, Berkeley, all speakers read and signed letters of permission both before and after their recording sessions.

systematic linguistic analysis, and in gaining insights into local language ideologies and linguistic norms.

Sociolinguistics has long been concerned with the problem of the observer's paradox, and Labovian sociolinguistic investigations (e.g. Labov 1972) have demonstrated that people will speak differently when aware of observation. However, Labovian variationist investigations fall into just one of the three social science research stances currently in use: *positivism*, delineated by its commitment to the study of observable phenomena. Cameron et al. describe two alternate stances: *relativism*, which does not accept the distinction between facts and personal valuations, and *realism*, where observation is seen to be inherently theory-dependent, but reality itself is theory-independent (1992:7ff). According to Rampton (1992), the sociolinguistic fieldworker must additionally choose an affiliation of either Labovian variationist theory or the Hymesian ethnography of communication. Milroy (1987a, 1987b) believes that it is possible to maintain affiliation with both schools of thought and she attempts to synthesize their methodologies in a systematic way. My own research was conducted from within the realist framework, and I fall into the Hymes and Gumperz camp, and firmly believe that an extended period (or better, extended periods) of participant observation is necessary for any kind of sociolinguistic research. Perhaps this methodological bias stems from my research interests – I am not, at least currently, a variationist concerned with interspeaker variation in the performance of microvariables. In this section in particular, I will be concerned only with *intraspeaker* variation at the lexical level, and with style shifting.

The usual goal of sociolinguistic investigation is to gather recordings of “natural” speech, which is to say, somehow accessing and recording what is generally referred to as the vernacular, as “untainted” by interactions with the fieldworker as possible. Although specific definitions vary, among sociolinguists it is generally agreed that the vernacular is “the

relatively homogenous, spontaneous speech reserved for intimate or casual situations...taken to reflect the most systematic form of the language acquired by the speaker, prior to any subsequent efforts at (hyper-) correction or style shifting...” (Poplack 1993: 252). Historically, sociolinguistic investigators have also used the term “vernacular” to refer to a low-prestige variety in contradistinction to a standard and high-prestige variant, associating the vernacular with social groups (e.g., African Americans) or with localities (e.g., Belfast English) (Milroy 1987b: 58). The commonality in these usages is found in (1) their assumption of a monolingual, although bidialectal, situation, and (2) the limitation of vernacular use to in-group and intimate contexts.

My fieldwork situation was bilingual rather than bidialectal, and without a ready binary distinction of standard vs. vernacular varieties. The majority of Tatar speakers who belong to the Tatar Social Club were between 8 and 14 years of age in 1990 when Tatarstan declared sovereignty, and have grown into adulthood in the transitional post-Soviet world. As was seen in sections 1.2 and 1.3, Russian remains the dominant language in all spheres of life, and is high prestige, but in the milieu of the Tatar Social Club, as well as among Tatar intelligentsia, the Tatar language, more precisely “pure” Tatar (*saf tatar tele*) and “literary” Tatar (*ädäbi tatar tele*), is also awarded high prestige. Additionally, as will be seen below, there is a continuum of linguistic performance for urban Tatar bilinguals, ranging from completely Russian (in both formal and informal registers) at one end of the cline to completely Tatar (in both formal and informal registers) at the other end, further complicating the designation of a single “unmarked” vernacular.

The goal of accessing the “vernacular” by means of “natural” speech not only relies upon a positivist stance, it also works from the assumption that one can avoid the effects of the fieldworker as a socially located person (but see criticisms by Cameron et al. 1992 and

Milroy 1987a and 1987b; see also Besnier 1994, Ochs 1998, Schilling-Estes 1998). While some linguistic fieldworkers (e.g., Baugh 1993, Besnier 1994) claim that informants can forget about recording and produce naturalistic speech, I found that in my own fieldwork, the presence of the language investigator, with or without recording materials, would often be enough in and of itself to precipitate “performed” speech rather than “unperformed.” The word “performance” has a variety of connotations in linguistic study; in this context performance speech is “associated with speakers’ attempting to display for others a certain language or language variety, whether their own or that of another speech community” (Schilling-Estes 1998: 53). Only after my return home from the field, and extensive review of my recordings and fieldnotes, did I realize that the linguistic performance and style shifting of the Tatar speakers I had observed and recorded could be best interpreted by taking into account two major factors: (1) the speaker’s assessment of my social role, particularly, in the broadest terms, as in-group or out-group member, and (2) my participant role in the speech event in question.

1.4.1 Observation and audience design

I will begin with an examination of the participant role of fieldworker, and reconsider the observer’s paradox using insights gained from Bell’s (1984) theory of audience design. This reconsideration will not attempt to predict the effect of the fieldworker upon speech production, but will allow for a more systematic accounting for the effect of observation and recording upon performance, and thus for a more nuanced and accurate linguistic analysis of data gathered during fieldwork.

The first sociolinguistic investigations conceptualized style shifting on a single continuum ranging from careful to casual speech. Labovian sociolinguistic interviews were

designed to elicit more- and less-careful styles, and topics introduced by the interviewer, including questions about childhood experiences and brushes with death, were meant to create contexts for casual speech, where attention paid to content would supersede attention paid to form, as well as awareness of the observation of the language investigator. While investigations of this sort attempted to use controlled style shifting as a research heuristic, the methodology used by the sociolinguistic interviewer to manipulate speaker styles came under critique for a variety of reasons, and in recent years sociolinguists have been focussing on style shifting as a naturalistic phenomenon.

Following Brown and Levinson (1979), Bell (1984) dismissed Labov's conception of style as varying according to the amount of attention paid to speech as an "impoverished" view. He proposed an alternate explanation, that of *audience design*, which holds as a basic tenet that "at all levels of language variability, people are responding primarily to other people. Speakers are designing their style for their audience" (197). Variation can be found in two dimensions: the social dimension, which is expressed as *interspeaker* variation, and the stylistic dimension, which is expressed as *intraspeaker* variation, and Bell posits that style is derived from the social, such that "intraspeaker variation is a response to interspeaker variation" (158). Variables such as topic and setting are seen to have less effect upon stylistic variation than audience, which is the "responsive, critical forum before whom the utterances are performed" (161): for empirical testing of this claim cf. Rickford and McNair-Knox (1994) and Lewis (2002), *inter alia*. In fact, Bell proposes that the stylistic effects of topic and setting are ultimately derived from the audiences with which they are associated. The audience design framework is generally held to be superior to the attention-to-speech continuum (which has fallen into disuse), since this single organizing principle can account for such diverse behavior as bilingual code-switching, politeness strategies, and care-taker

speech; the careful or self-conscious use of the “vernacular;” and peer group maintenance of “vernacular” norms (Milroy 1987b: 179).

Audience composition is seen to be heterogeneous, and refining upon Goffman’s (1981) “participant framework,” Bell offers five participant roles for any given speech situation, which are as follows:

1. Speaker – uses the 1st person,
2. Addressee – addressed in 2nd person,
3. Auditor – referred to in 3rd person,
4. Overhearer – unratified to participate in the speech event, not addressed or referred to,
5. Eavesdropper – both unratified and unknown.

Each step down on the list increases the distance from the speaker, sometimes physically.

Figure 1 below summarizes the attributes of audience roles in a speech event. A fieldworker can participate in any of these audience roles.

Figure 1. Hierarchy of attributes and audience roles (adapted from Bell 1984)

	<i>Known</i>	<i>Ratified</i>	<i>Addressed</i>	<i>Person</i>
Addressee	+	+	+	2 nd
Auditor	+	+	-	3 rd
Overhearer	+	-	-	n.a. (unratified)
Eavesdropper	-	-	-	n.a. (unratified)

Audience roles are assigned by the speaker, and will have different levels of salience for the speaker’s style design – for example, auditor effect is usually lower than addressee effect. Style shifting can result in convergence to or divergence from the audience: convergence is seen as accommodation, and an expression of what Brown and Gilman (1960) call “solidarity,” while divergence is interpreted as “referee design,” which marks the speaker as a member of a social group not present in the speech situation – but referred to by his or her divergent style – and marks the audience (most commonly an addressee) as a non-member of the referred-to group.

Therefore, one must take into account both the fieldworker's participant role in a speech event and her or his position as a socially located being. Milroy, arguing for the necessity of studying language in context in order to discover the "total linguistic repertoire," was the first to propose that both the social networks of speakers being studied and the location of the fieldworker within those social networks need to be taken into consideration, warning that it is "unwise to underestimate the importance of a careful choice of fieldwork method; for...this choice has considerable influence both on the kind of language available for analysis, and on the ultimate analytic procedure" (1987a: 2). Milroy's sociolinguistic fieldwork in Belfast was conducted from the social position of "a friend of a friend," which is to say, she gained entrée to local networks through mutual acquaintances, and then expanded her network connections in the community by means of introductions made by the people with whom she had first become acquainted. Milroy felt that this social position of "friend of a friend" meant that the community simultaneously perceived her as both insider and outsider, and that this ambiguous social perception allowed her to collect data from both perspectives. Milroy, Li, and Moffat (1991:288) suggest that audience design theory calls for "both field methods which reduce the prominence of the investigator, and for analytic procedures which account for his/her interactional role." When writing about sociolinguistic investigations, linguists seem to have as a prototype the urban, monolingual context with a local, non-standard dialect, where the fieldworker is a native speaker of the language being investigated,³⁶ but Milroy, Li, and Moffat expanded their analysis to include sociolinguistic fieldwork in bilingual communities. Li, ethnically Chinese and long-integrated

³⁶ This is in contrast to linguist anthropologists, whose prototypical situation seems to be a rural, isolated monolingual village where the fieldworker must acquire the language being investigated.

into the community he studied, turned out to elicit the same speaker design as community members of his age and sex, such that younger members of the community would not use English when he was an addressee. Moffat, studying acquisition of English by Panjabi kindergarten children, and a monolingual English speaker, had the same effect on language switching as the children's teacher: their presence as auditors would cause a shift to English. Drawing on these experiences, Milroy et al. suggested that by finding the social role within the experience of the speakers being studied that elicited the same linguistic performance as the fieldworker, one could more systematically account for the effect of the observer on speech production: Li was treated like other adult in-group members and Moffat was treated like other adult out-group members. However, Milroy et al.'s taxonomy of fieldwork situations falls short, since it works on the presumption that an "outsider" fieldworker will be a member of the majority social group, here, a white British monolingual English speaker, and thus easily classifiable by minority group members. In fact, there are two kinds of "outsider" fieldworkers: those who are members of the society at large in which the local (minority) community is situated, and those who are completely alien – neither a member of the dominant group nor a member of the minority group, and a native speaker of neither language. For many speakers in the community being studied, contact and interactions with this fieldworker may not be comparable to interactions with any other person in their experience. Therefore, finding a local parallel in order to account for the fieldworker's effect on speaker performance in the same way simply may not be a viable option. On the other hand, this uniquely ambiguous social positioning can be utilized by the canny fieldworker to facilitate his or her investigation.

For example, as happens with so many other investigators of minority languages, particularly those undergoing a multi-generational language shift, my foreignness, combined

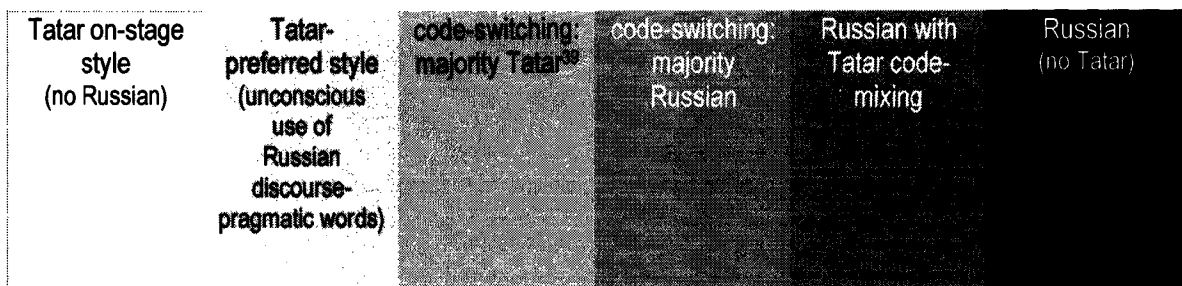
with my statement of purpose – that I was in Tatarstan to learn and study Tatar language and culture – was met with surprise, and would elicit commentary from all and sundry. Both my stated intention to study and learn Tatar and my attempts to speak Tatar would elicit commentary on my performance, on the linguistic performance of other Tatars, on the Tatar language itself, and on the sociolinguistic and cultural context of post-Soviet Tatarstan. It quickly became clear that Tatar speakers felt responsible for presenting me with the best possible Tatar, both so I could have appropriate models for learning, and so I could represent the language well in my research, and record a literary (*ädäbi*) and beautiful (*matır*) version of the language for posterity. My lack of Russian phonetic and phonological interference in Tatar would cause young urban speakers to reflect disparagingly on their own Russian-accented speech, and young person after young person upon meeting me would offer to take me “back to the village,” where I could hear “real” village Tatar spoken – only Tatar, all day, every day, Tatar that was “purer” than the urban dialects, which had been tainted by Russian.³⁷ Speakers would, unasked, happily rate the style, presentation, and competence of other speakers, suggesting some as appropriate models for emulation and study, and dismissing others. As my Tatar competence improved, I found that this high level of language awareness, with its stated ideals of “pure” Tatar and “literary” Tatar, was not merely provoked by the presence of the language learner and investigator. This “discourse of purity” and constant awareness of the level of Russification of Tatar (mostly on the phonetic and lexical levels), combined with the idea of “saving” the Tatar language from both misuse and disuse, was not only presented to me but also found in discourse in newspaper articles

³⁷ In point of fact, young people in villages do not speak only Tatar all day every day, as became immediately apparent upon my visits.

and opinion pieces, on television, on internet bulletin boards, and on the radio. (For an in-depth examination of the Tatar “discourse of purity”, see Chapter 2.)

The young speakers whom I studied gave me only limited access to their range of performance styles, particularly at the beginning of our acquaintanceship, but over time, as their familiarity with me increased and my positioning within the community became increasingly in-group (and my skills in overhearing and eavesdropping on public conversations improved), I eventually gained access to a wide range of styles. What I found was that urban bilingual Tatars had a continuum of linguistic performance that was related to language mixing, as shown in Figure 2, below.

*Figure 2. Cline of language mixing and performance styles for urban Tatar bilinguals*³⁸



Tatar on-stage style is often found in the public sphere in formal registers, particularly when aiming for a high literary standard, but can also be found in informal register. In this style,

³⁸I would like to acknowledge up front that these styles are arranged according to a single stylistic variable, the level of use of Russian, even though there are other markers that differentiate, for example, Tatar on-stage style from Tatar-preferred style. Additionally, linguistic performance within each style can vary, e.g., Tatar on-stage style can be both in formal register and in informal register. And within formal register of Tatar on-stage style there are further nuances, such that speeches given at a literary tribute evening are different from the on-stage performance of the master of ceremonies at a concert. I in no way wish to suggest that this single variable, level of Russification, is the sole differentiating marker of Tatar style and performance.

³⁹ My delineation of code-switching into two styles, ‘majority Tatar’ and ‘majority Russian’ is similar but not identical to what Myers-Scotton would delineate as Tatar as ‘matrix language’ or Russian as ‘matrix language.’ For more on the similarities and differences between our typologies, see section 3.2 of Chapter 3.

speakers will “de-Russify” their Tatar to the best of their ability. Performances in Tatar on-stage style will generally take place at Tatar cultural events, in radio and television interviews, in political speeches, and in public presentations or comments at the Tatar Social Club. Private use of Tatar on-stage style seems to only occur in conversations with or for investigators of Tatar language and culture.

Tatar-preferred style is always in informal register, and is found in private conversations where Tatar has been in some way established as the preferred language of communication: for example, this is the main style of intergenerational family communication in Tatar-speaking homes, both urban and rural. For intragenerational family interactions, generally Tatar-preferred style will be used by parents, while siblings will interact in one of the code-switching styles, or even in Russian with Tatar code-mixing. In Tatar-preferred style, the level of language awareness and verbal hygiene is lower than in Tatar on-stage style, and although speakers believe themselves to be speaking “pure” Tatar, they actually are unconsciously using Russian discourse-pragmatic words. These code-mixed words – not standard borrowings, and mostly function words, adverbs of manner, discourse particles and interjections – both structure and comment upon discourse. (See Chapter Three for a detailed examination of Tatar-preferred style and the code-mixing of Russian discourse-pragmatic words).

This cline of language-mixing that I found in urban Tatarstan parallels the findings of Eliás-Olivares, who studied Spanish-English bilinguals in Austin, Texas of the 1970s (Eliás-Olivares 1976, cited in Pfaff 1979), and found this style continuum:

Standard Spanish – Popular Spanish – Español Mixtoreado – Caló – Chicano English – Standard English

Standard Spanish is Standard Mexican Spanish, while Popular Spanish is the non-standard local version, “characterized by a number of phonological changes and morphological regularizations” (Pfaff 1979: 292). The variety here called Español Mixtoreado, also known as Pocho or Tex-Mex, is characterized by its high level of Anglicization, including borrowings and calques. The style known as Caló is used mostly by adolescents, and characterized by a high level of Anglicisms mixed with Spanish slang. Older bilinguals in Austin “look down on Español Mixtoreado and code-switching. For them, ‘good Spanish’ means the Spanish of Mexico, and, in particular, the avoidance of Anglicisms” (Elias-Olivares 1976:152 in Pfaff 1979). This desire for language purity parallels the one found among Tatars, many of whom stigmatize code-switching styles, which are by no means limited to younger generations.

1.4.2 The role of the language investigator in style shifting

As my fieldwork progressed, I was able to gain access almost all of the styles shown in the continuum above, and managed to observe much intraspeaker variation in linguistic performance. However, it was only later that I realized that by paying careful attention to the placement of the language investigator (or out-group language learner) in the participant role framework of audience design, I was able to discern patterns in style shifting that showed that a speaker’s choice of style was dependent in part upon the audience role assigned to the fieldworker participating in a speech event.

I will now briefly present the style shifting of one Tatar bilingual, who I will call Galimä,⁴⁰ a Tatar philologist with whom I had both a professional and personal relationship. I observed Galimä in a variety of situations, locations, and speech events: at her house with

her family and with guests, teaching a class of lycée students, lecturing in front of college students, lecturing in front of low-level government workers, conversing with colleagues at both the lycée and university, conferring with and advising younger Tatar philologists, running errands at the bank and post office, conversing on the street with friends and acquaintances, chaperoning a field trip of lycée students, conversing with parents of her students, and more. I believe that for Galimä my social role remained reasonably constant – I was non-peer (15 years younger), outsider, language learner and language investigator. Although our relations were quite warm, and we discussed personal matters frequently, she would only address me using the formal version of “you” (*sez*) and never once used the informal “you” (*sin*).

If we exclude Galimä’s professional lectures and look at only her spontaneous speech production, we can see the relationship between her style shifting and my placement within the participant framework, including my salience and ratification in the speech event – I was at various times addressee, auditor, overhearer, and eavesdropper. Additionally, we will see that my presence would sometimes cause Galimä to engage in metacommentary upon style shifting and her linguistic performance.

(1) In dyadic conversations (where I was both a speaker and an addressee) in either a professional or a personal context, Galimä was always in Tatar on-stage style. As with many other Tatars who felt responsible for helping me in my Tatar language learning and investigations, she would choose ideology over communication, and *never* used Russian with me, opting for a simplified Tatar to explain words or phrases I hadn’t understood, when a single Russian word would have sufficed.

⁴⁰ A pseudonym I have chosen for its meaning of ‘scholar,’ ‘knowledgeable one.’

(2) Galimä would use Tatar on-stage style with colleagues, friends, and acquaintances when ratifying them as participants in conversations in which I would be speaker, addressee, or auditor. She would do this even with speakers whom she knew to have limited Tatar competence. This was clearly a self-conscious “performance” of Tatar, and arrangement of performance by others, one that was staged for my benefit. Very often it was only Galimä’s participation in the conversation that kept it in Tatar on-stage style; if she would leave, speakers would frequently ask in Tatar if I knew Russian, and when I answered in the affirmative, would either switch to Tatar-free Russian or would code-switch with Russian as the majority language.

(3) If I was an auditor of a conversation with family and friends, Galimä would speak in Tatar on-stage style, with no Russian whatsoever. For example, if we were drinking tea in her kitchen and talking, and her son came in and asked a question, she would use only Tatar with him. Additionally, if he used any Russian, she would upbraid him and tell him to speak in Tatar only. However, if I was not an auditor but rather an overhearer, and thus unratified as a conversation participant, Galimä would code-switch with Tatar as the majority language. For example, one time I arrived early for a meeting with Galimä that was going to take place in a classroom in the lycée where she taught, and entered the room while she was still consulting with her younger cousin, advising her on how to best teach Tatar periphrastic verbs. Galimä saw me enter the room, and nodded in acknowledgement – I sat at the opposite end of a long table, read a newspaper, and surreptitiously listened as the two spoke mostly in Tatar, but with code-switching into Russian of higher-level grammatical constituents and code-mixing of isolated words. As soon as their meeting was over, I became a ratified conversation participant, and when Galimä introduced me to her cousin, she switched into Tatar on-stage style, with no Russian, and her cousin followed suit.

(4) Galimä would speak Russian with those bilingual Tatars with whom the Russian is the usual language of interaction (due both to sociolinguistic conventions and Tatar's functional domains), but if I was an auditor of one of these Russian-language interactions, my auditing would cause linguistic metacommentary on Galimä's part. For example, one time en route to a class where Galimä was going to introduce me to some potential subjects, we stopped off at the university post office. Galimä and I were speaking Tatar up to the moment of the transaction, and the postal worker was speaking on the phone in code-switched Tatar and Russian, but the transaction itself was conducted in Russian, as is typical. I too had business that was also conducted in Russian. When the transaction had concluded and we turned away to leave, resuming our Tatar-only conversation, Galimä said to me, "I don't know why I speak Russian with that woman. I've been coming here for years. She knows that I speak Tatar, and I know that she speaks Tatar. So why do we speak in Russian?" I don't want to say that my presence brought this fact to Galimä's attention for the first time, but I believe that my conversational auditing did trigger the meta-commentary. Perhaps this is because of the conflict that my auditing caused in Galimä's style-shifting. Recall that Galimä's usual behavior when I was going to be a ratified conversation participant was to use Tatar whenever possible with anyone whom she knew to be a Tatar speaker, regardless of their competence. However, the local linguistic norms required her to shift to Tatar-free Russian when transacting post office business, and this requirement seems to have superseded the style-shifting patterns that were based on my presence. Perhaps this conflict in linguistic presentation of identity – culturally competent citizen on the one hand, and practitioner of verbal hygiene and performer of pure and literary Tatar and the other – is what caused her explicitly stated dissatisfaction with her linguistic performance.

(5) When conversing with monolingual Russians, both ethnically Russian and Tatar, Galimä would speak Russian only, regardless of my participation role. This would sometimes lead to an interesting phenomenon – a sort of Russian hangover, where Galimä would be “out of phase” in her style shifting, such that after the Russian-language conversation had ended, she would return to our conversation and address me in Russian. However, this Russian performance would only last for one conversational turn, because regardless of the language of my response, Galimä would become immediately aware of her “inappropriate” style. I could respond in Russian, or I could respond in Tatar, but her response, always in Tatar, would be the same: “Why am I speaking Russian with you? We don’t speak Russian together. Let’s speak Tatar.” And we would continue our conversation with her in Tatar on-stage style.

In summary, we can see that although my social role remained constant with Galimä, my various participation roles in speech events seem to have influenced both her style shifting and her awareness of style shifting. Perhaps this can be seen more clearly in Figure 3, below.

Figure 3. Effect of changing participant roles of fieldworker on Galimä’s speech

<i>Fieldworker role</i>	<i>Speech event participants</i>	<i>Situation</i>	<i>Galimä’s speech style</i>
Speaker/addressee	friends; family; colleagues; acquaintances; no other participants	private conversation	Tatar on-stage style
Auditor	friends; family; colleagues; acquaintances	private conversation	Tatar on-stage style
Overhearer	friends; family; colleagues; acquaintances	private conversation	Code-switching: Tatar as majority language
Auditor	service personnel	business transaction	Russian with post-transaction metacommentary
Auditor	Russian monolinguals	all	Russian with periodic “Russian hangover” followed by metacommentary

Instead of labeling any of this behavior as simply the “observer’s paradox,” or “unnatural” or “performed” speech due to the presence of the fieldworker, I believe analyzing the linguistic performance in question from within the framework of audience design theory gives a more nuanced and systematic account, and allows us to unpack the various types of performance, which cannot simply be assigned to one or the other pole of the binary opposition of natural and unnatural. Because of my social role as language investigator, I had limited access to Galimä’s range of performance styles: I was only able to observe three out of the six styles found in the continuum shown above in Figure 2. I believe that this is simply a limitation of the fieldwork situation: Galimä’s construction of her sociolinguistic identity only allowed three of these styles to be appropriate in my presence – truly “unnatural” behavior would have been inappropriate linguistic performance on her part.

1.4.3 The role of the language investigator in recording speech

The various performance styles that I have just described were produced in the presence of a fieldworker who was not explicitly in her role as fieldworker at the time of the speech events in question, and who was without any obvious note-taking or recording equipment. Now I would like to turn to the recorded speech event, and briefly examine this too from within the framework of audience design. A recorded private-domain speech event, regardless of the physical presence or absence of the fieldworker, is an atypical, even extraordinary situation that is not classifiable within Bell’s ordinary hierarchy of audience roles. And here is why: the recording equipment, previously analyzed by some sociolinguists as a participant itself in the speech event, actually represents an end-listener or listeners whose identity is not known at the time of the speech event. This means that the speech event participant represented by

the recording equipment is simultaneously *ratified* (providing that permission to record has been requested and granted) and *unknown* – a participant role that is not analogous to any found in Bell’s framework because it is unique to the experience of being investigated by a fieldworker of some sort, linguistic or otherwise. In Figure 4 below I have added the end-listener of recorded speech to the audience role hierarchy, so that it can be easily compared with the other standard audience roles. The “strange” or “unnatural” behavior of recorded speech event participants can thus, in part, be interpreted as speakers trying to grapple with a participant role they have never dealt with before, that of the unknown eavesdropper who is nonetheless ratified. Perhaps what is unnatural is not so much the behavior and performance of speakers as the uniquely conflicting audience role attributes of the recording’s end listener, for whom speech must be designed, but whose social role and identity remains cloaked in mystery (or at least unclear).

Figure 4. Adjusted hierarchy of attributes and audience roles for a recorded speech event

	<i>Known</i>	<i>Ratified</i>	<i>Addressed</i>	<i>Person</i>
Addressee	+	+	+	2 nd
Auditor	+	+	-	3 rd
Overhearer	+	-	-	n.a. (unratified)
Eavesdropper	-	-	-	n.a. (unratified)
End-listener of recorded speech	-	+	-/+	2 nd /3 rd /n.a.

“Unnatural” or “self-conscious” behavior seen to be a result of the observer’s paradox can thus be more appropriately accounted for through this refinement of Bell’s audience roles. Recall that according to Bell, speakers design their speech with audience members in mind, and those participants whose audience roles are higher up on the hierarchy (as seen in Figure 1) will have a greater effect on the linguistic performance of the speaker. However, research on some non-Western speech communities has shown that auditor effect can be equal in

strength to addressee effect: for example, Jahangari's work on Tehranian Persian (1980, cited in Bell 1984: 175) showed that having high status auditors would cause speakers to use only formal second person address forms, even for close friends who would have been addressed using the informal pronoun in the absence of such an auditor. I submit that the unnatural audience attributes of the end-listener of recorded speech, highlighted by the act of recording and the presence of recording equipment, can cause this audience role to be of primary salience and focus, regardless of the fieldworker's actual participant role in the speech event. The fieldworker may be entirely absent and yet still the most salient speech event participant. By understanding the identity and traditional audience role assigned to this end-listener by the speakers being recorded, it is possible to more systematically account for her or his effect on recorded speech. The end-listener may be seen as entirely congruent with the fieldworker, and the fieldworker alone; can be perceived as the fieldworker in combination with other language investigators; and can be perceived as a person or persons completely unknown to the speaker.

Before I give a few examples of end-listener roles assigned by my Tatar friends and acquaintances who were being recorded, I think it is important to note that my fieldwork took place in a society that has been grappling with observation and unratified eavesdroppers for many years – to wit, the surveillance of the KGB (which has retained its name in Tatarstan, although it is now referred to as the *FSB* elsewhere). This constant awareness, particularly for the more politically active members of the Tatar Social Club, who see themselves as opposed to the government at both the republic level (Tatarstan) and the federal level (Russian Federation), certainly affected their speech production. Although I managed to convince most members of the social network I was studying that I was not a

spy working for the CIA⁴¹ – a question that was put to me point blank in my fourth week onsite – there was always the possibility that my recorded materials and notes would be confiscated, either in Russia or after my return home.⁴²

Some of the more politically involved community members, particularly those who were well aware that they were under surveillance by the authorities – surveillance that predated my arrival – seemed to not merely view the end listener’s identity as just unclear, but also quite likely to be an unknown, and unfriendly, intelligence officer. Their verbal hygiene involved cleaning up their speech in order to speak about only the most inoffensive and unimportant of topics (thus limiting any access I might have had to the nature of the performance of political discourse), or they would simply shut off the recording equipment if they were joining a conversation that was in progress and they didn’t feel like watching their tongues.⁴³

For some other Tatar speakers, the end-listener was presumed to be the language investigator, and the language investigator only. In this scenario, the end-listener took on the attributes of addressee, despite physical absence at the speech event. For example, in three separate recorded conversations, two of which were recorded in my kitchen while I was out

⁴¹ Apparently some members of speech communities that have been under study will only deem their language investigator trustworthy (and not a spy) decades after the original fieldwork (Penny Eckert, personal communication).

⁴² This possibility was thrown into higher relief by the Russian-American “spy wars” of 2000-2001, with “embassy workers” expelled from both Moscow and Washington, D.C., and a young American scholar imprisoned on both drug possession and espionage charges (although the espionage charges were later dropped).

⁴³ I should also note that some of these community members believed both my apartment and telephone to be bugged, and that either their e-mail account or my e-mail account was being read (I have it on good authority that mine was not). When speaking with me on the telephone, they would sometimes say, “This is not a phone conversation, I’ll tell you about it later.” Or, they would pull me out onto my balcony, presumably out of range of any bugs inside the apartment, to whisper anything of important they had to tell me, for example, about their knowledge of KGB surveillance of me.

of earshot in the living room, and a third which was recorded in one speaker's home on a tape recorder I had lent her, conversation participants performed for me in the most standard sense of the word, singing Tatar songs directly into the recording equipment (one performance is even in three-part harmony). Several of the participants – none of whom had access to the performances of the other recording sessions – later volunteered the information that they had sung for my sake, as a souvenir (the expression in Tatar is translated literally as “for the memory”) so that later on when I was back home in America and far away, I would have these nice songs to listen to, and think of them fondly. We can extrapolate from this behavior and subsequent metacommentary that the speech events being recorded, into which the singing was embedded, were performed with consciousness of my eventual review of the recorded material and subsequent linguistic analysis. These people were speaking “good Tatar” both for me and for posterity, and their level of verbal hygiene remained quite high – they stayed well on the Tatar side of the cline of language mixing, never moving further towards Russian than Tatar-preferred style, and remaining for the most part in Tatar on-stage style. This parallels the behavior of one speaker of Ocracoke English studied by Schilling-Estes' (1998), whose style shifting into “performance” speech was triggered not by a change in audience, but by a change in the perceptual salience of the audience make-up, i.e., the language investigator was suddenly made salient through attention to technical details of recording: changing a tape, flipping it over, or orally marking the tape with the name of the interviewee and date.

1.4.4 The value of performance speech

Bauman (1975, 1977) has argued that in addition to “vernacular” speech, “performed” speech is available and meaningful for analysis, yet it remains understudied by sociolinguists

and linguistic anthropologists. Self-conscious or cleaned-up speech styles, especially when one is able to compare them with other speech styles and has access to intraspeaker variation, can give insights into local linguistic norms and language ideologies (for work on minority-language language ideologies, see Woolard 1989, Hill 1998, and Kroskrity 1998, in particular), and lead to unexpected and useful avenues of linguistic inquiry.

Ochs (1978) describes how her initial attempts to gather the spontaneous speech of Samoan children and their caretakers were stymied by the Samoans' awareness of both her location – physical and social – and her recording equipment. The Samoans' use of “on-stage” and “off-stage” registers (Goffman 1963), in local terms known as “good speech” and “bad speech”, showed systematic patterning when understood within the local framework of spatial organization, where the front and center regions of a house were associated with more-controlled demeanor and activities and the back and peripheral regions were associated with less-controlled demeanor and activities, including their linguistic counterparts of “good” and “bad” speech. Ochs' initial placement of her mat and recording equipment at the front of the house and her continued presence on the mat during her data-gathering time located her physically in the on-stage and more-controlled zone, and located her socially as a high ranking guest and foreigner. Once she understood this system of spatial contrasts, she was able to “exploit [her] knowledge of space and language to elicit a wide range of speech styles from children and caretakers” (1978: 2). Additionally, Ochs was motivated by this experience to study children's acquisition of language and culture in the Samoan context and the systematic relation between linguistic forms and social situations, turning what could easily have been dismissed as the observer's paradox and useless data into a fruitful line of inquiry.

For Schilling-Estes, performance speech in Ocracoke English turned out to give useful insights into sociolinguistic variationist theory. She found that her informant's raising of the nucleus of the diphthong /ay/ in performance speech matched his patterns of /ay/ raising in other styles, with one major exception: while in non-performance speech, he would raise /ay/ to its highest level before voiceless consonants, in performance speech he would raise /ay/ to its highest level before voiced consonants, the raising environment that is most salient to speakers of the Ocracoke dialect. From this, Schilling-Estes concluded that when speakers "attempt to 'put on' a dialect for an audience, they have available to them only those features they can perceive... Thus, through examining performance speech, we can gain insight into which aspects of linguistic production are most salient to the performer" (1998: 64). In my own research, performance speech was useful in showing what aspects of Russian influence are most salient to Tatar bilinguals, and thus available for cleansing when verbal hygiene is at its highest level. I found that the most salient aspects of Russification were code-switching (of high-level grammatical constituents) and the use of Russian loanwords that are either not standard borrowings or have a post-*perestroika* (and thus pre-Revolutionary) Arabic equivalent whose use is being encouraged, e.g., using the Russian word *vlast'* rather than Arabic *khakimiyat* to express the concept of 'ruling power.' In a study of style shifting in East Sutherland Gaelic, a moribund Gaelic dialect used only in the private sphere, Dorian (1994) describes similar behavior, where performed speech, or more formal narrative style, was marked in part by the removal of Anglicisms. She writes, "When I first began to gather tape-recorded texts from each of the East Sutherland fisherfolk villages, it was quickly apparent that those speakers who had the lexical latitude to do so made an effort to minimize their reliance on loanwords from English" (ibid. 222). Dorian found performance speech useful not only for highlighting what dominant language influence was

salient to local Gaelic speakers, but also the varying degrees of competence of her informants.

Tatar on-stage style, particularly when combined with metacommentary and speaker evaluations by members of the speech community, was absolutely indispensable in my understanding of a variety of local norms and practices. “Performance” speech pointed the way to the “discourse of purity” that was the most significant of the language ideologies that I found, and exemplified speaker’s attempts to produce “pure” Tatar. It demonstrated verbal hygiene in practice, and showed me who were the community’s standards keepers. When speakers remained in Tatar on-stage style rather than accommodating their uncomprehending audience by switching into Russian, it alerted me to the existence of Tatar speakers who choose ideology over communication, for whom out-group referee design supersedes accommodation and convergence. Tatar on-stage style, when produced by less-competent Tatar speakers who feel compelled to refrain from using any Russian at all, can provide excellent examples of morphosyntactic interference. For fully competent speakers, a comparison of their linguistic performance in Tatar on-stage style and in Tatar-preferred style gave evidence for what appeared to be two different underlying competences and grammars: using only one of these styles for a grammatical description would lead to an incomplete and erroneous analysis of the level of Russian influence on Tatar grammar. Finally, placing “performance” speech within a continuum of language-mixing styles gives insight into the identities and roles that are associated with certain kinds of styles, and how choosing styles when constructing identity within this context may play a role in the stylistic and domain shrinkage of a language, and thus play a role in language shift.

1.4.5 Conclusions

Style shifting can be analyzed as either reactive or initiative (sometimes called “proactive”), and Bell constructs this reaction-initiation distinction as a continuum rather than as a dichotomy (1984: 184). His interpretation posits initiative style shift as used by a speaker “to redefine the nature of the interaction rather than to respond to different situations” (ibid. 183). Bell’s proposed interpretation of initiative style-shifting parallels Blom and Gumperz’s (1972) “metaphorical” code-switching, e.g., where speakers mark a change in formality (switching from formal pronouns to informal, or vice versa), or mark a change in attitude toward the topic. Schilling-Estes suggests that we conceptualize style shifting “in terms of roles (both real and metaphorical) that conversation participants play with respect to one another during a given interaction” – these roles are simultaneously permanent social relations and “individually based expressions of identity” (1998: 72). A single group of conversation participants can be seen as containing a variety of role relationships, and the speech of an individual as containing a variety of voices, as suggested by the Bakhtinian conception of heteroglossia (1981), where speech styles and registers are evocative of various cultural and ideological positionings – using the different “voices” associated with a social group, cultural position, or ideological stance allows the speaker to shift among roles evoked by these “voices.” In her early critique of the Labovian sociolinguistic interview as a method for collecting spontaneous speech, Wolfson called into question the idea of a “single, absolute entity answering to the notion of natural/casual speech” (1976: 202). And indeed, from the perspective of Bakhtinian heteroglossia, all speech can be seen as performative in some way, from high-prestige standard styles to local, in-group, non-standard styles. Studies of supposedly “unperformed” vernacular styles that have non-standard norms rigorously

maintained by standards keepers can be interpreted as coherent with audience design theory. Community members must use non-standard variants in in-group styles, and failure to use these variants can result in punishment ranging from teasing to actual physical beatings (cf. Milroy 1987a). Not one of the styles that I observed among urban Tatar bilinguals can be regarded as more natural or unmarked than any of the others: they are all designed to be appropriate for the audience, topic, and setting of the speech event in which they are produced.

Interpreting all speech as performative, and thus rendering futile the search for a natural unmarked vernacular, has ramifications for linguistic fieldwork and analysis. The first is that no speech should be dismissed as unnatural and thus unusable: the distinction of natural and unnatural, usable and unusable is not a viable one. Instead, attention should be paid to the contextualization of the speech event: the language investigator should attempt to discover who is performing, what audience they are designing their speech for, which participant role in the audience composition is most salient or has the greatest effect, and finally, what role the speaker is performing, with an understanding that which identity is being presented or constructed through language may change in the course of a speech event, triggered by factors other than change in audience composition. This is a significant amount of information and contextualization for a single speech event, and clearly requires an intimate knowledge of local norms, and of the speaker and audience involved. Therefore, long-term participant observation can be seen as an absolute necessity for any sort of nuanced sociolinguistic analysis. Second, performance speech can be utilized for linguistic analysis in a variety of ways, for example, to elucidate verbal hygiene patterns, or to demonstrate which elements of a language or dialect are most salient to its speakers. Third, the fieldworker should try to gain access to as many styles and registers as possible – with an

understanding that not all speakers will produce all styles for a language investigator – by exploiting her or his simultaneous insider and outsider status, as well as her or his changing status over time. Additionally, it should be understood that the fieldworker will have access to some styles when participating and observing, but will be denied access to them with the introduction of recording equipment to the scene of the speech event. Varying one's recording methodology – recording when the fieldworker is both present and absent, recording in different settings, and having community members record for you – can increase access to various styles, particularly those that are in-group only, and reduce the salience of the “unnatural” audience role of the unknown yet ratified end-listener.

To conclude, I have briefly addressed just one of the difficulties of sociolinguistic fieldwork, the gathering and contextualizing of spontaneous speech. I have attempted to demonstrate that by framing the observer's paradox within audience design theory, it is possible to more systematically account for and analyze the effect of the fieldworker as a socially located being, one who holds a unique social position, simultaneously insider and outsider. This ambiguity of role assignment can be exploited by the canny fieldworker in order to access a range of styles and registers. By following other sociolinguists and linguistic anthropologists in concluding that all speech in fact may be regarded as self-conscious, performed, and identity-constructing, I have once again called into question the idea of a single, unmarked, unperformed vernacular, the access to which is supposedly blocked by the observer's paradox. Finally, I have shown that a “performance” or “on-stage” style can be just as useful for linguistic and sociolinguistic analysis as other styles. By maintaining what Hyman (2001) calls the “fieldworker mentality,” which keeps the language investigator open to unanticipated phenomena, and by working from a realist research stance – in which reality is seen to be theory-independent, such that current linguistic theories may prove to be

insufficient as organizing principles for behavior and speech production – the fieldworker may find that performance speech will lead to both unexpected and unexpectedly fruitful avenues of linguistic inquiry.

Chapter Two

Linguistic purism and the “de-Russification” of Tatar

2.1 The Tatar “discourse of purity”

One of the major insights of 20th century studies in linguistics and the philosophy of language is that speaking is an act, and that as an act it is therefore comparable to other social activities: language does not just reflect and comment upon social reality, it also constructs it. As Silverstein (1979: 194) notes, due to the work of Wittgenstein, Austin, and Hymes, among others¹ “it has become clearer that people not only speak about, or refer to, the world ‘out there’ – outside of language – they also presuppose (or reflect) and create (or fashion) a good deal of social reality by the very activity of using language” (see also Gal 1989: 347, Blommaert 1999: 6). Language can thus be analyzed as multifunctional, both indexical and constitutive of social structure (Irvine 1989). The study of language as a practice and its relationship to language as a system is a major component of present-day linguistic anthropology and sociolinguistics. The links between language and the material world are well-studied, and sociolinguists have demonstrated correlations between identity (gender, age, ethnicity, class, etc.) and linguistic performance: these correlations are also referred to as “reflection” and “indexicality” (Gal 1989: 347). However, “correlational” sociolinguistic studies have come under critique by linguistic anthropologists, who contend that they do not take human agency into consideration, and thus “gloss over the actual motivating force of linguistic change, which often lies in social evaluations of language”

¹ Gal (1989) adds Kenneth Burke and Roman Jakobson.

(Woolard and Schieffelin 1994: 70). Studies of language as a practice argue that “in maintaining or changing local power relations, talk unites structure and agency over time” (Gal 1989: 347). Not only are languages “socioculturally and historically anchored phenomena,” discourse can be seen as “the crucial symbolic resource onto which people project their interests, around which they can construct alliances, on and through which they exercise power” (Blommaert 1999: 6-7).

One way that linguistic anthropologists integrate human agency and socio-cultural grounding into studies of language as a practice is through the investigation of language ideologies, which can be understood as the “mediating link between social structures and forms of talk” and “rooted in or responsive to the experience of a particular social position” (Woolard and Schieffelin 1994: 55, 58). The term *ideology* has been used in two distinct ways: neutrally, to describe cultural systems of ideas about social and linguistic relationships, and critically, where ideologies are seen as having both social and cognitive functions, and thus the ability to distort (ibid. 57): what is particularly interesting about the study of language ideologies is that they are located at the intersection of these two types. Traditionally there have been three genres of scholarly discussion that invoke language ideology: (1) studies of contact between languages and language varieties, (2) historiographies of linguistics and public discourse on language, and (3) studies of the relationship between language ideologies and linguistic structure (ibid. 55-6). This dissertation falls into the first and third categories, and in particular, the goal of this chapter is to make explicit the relationship between the most prominent language ideology found in post-Soviet Tatar discourse, what I call the Tatar “discourse of purity,” and the structure of the Tatar language as spoken in post-Soviet Tatarstan, including structural innovations (some of which which will be examined in depth in Chapter Three). Language ideologies, which are analyzed both through examinations of

explicit discourse and through examinations of implicit assumptions extrapolated from both social discourse and social practice, are never simply about language. They are, “among many other things, about the construction and legitimation of power, the production of social relations of sameness and difference, and the creation of cultural stereotypes about types of speakers and social groups” (Spitulnik 1998: 164). Ideologies, including language ideologies, are reproduced by practices, practices that can be institutional, semi-institutional, and everyday, found in governmental and educational institutions, advertisements, the media, literature, art, and more (Blommaert 1999: 10). “These reproduction practices may result...in *normalization*, i.e., a hegemonic pattern in which the ideological claims are perceived as “normal” ways of thinking and acting...There is no public opinion, no social consensus which can be detached from real processes of hegemonization” (ibid. 10-11). However, Gramsci’s conception of “cultural hegemony,” which “attempts to explain how and why subordinate groups accept as legitimate the power of the dominant,” implies “the possibility of resistance, or subversive practices and contradictory consciousness among subordinate groups” (Gal 1989: 348). The Tatar discourse of purity to be examined here can be defined as a “counter-hegemonic discourse.”

As was seen in Chapter One, in post-Soviet Tatarstan there is an asymmetry between the use and usefulness of Tatar and Russian: Russian is the “normal,” “unmarked” language that is used in all functional domains, while Tatar usage is “marked,” limited, and “particularized” (for more on the particularization of Tatar, see the work of Helen Faller, e.g., Faller 2001 and Faller 2003). The particularization of Tatar and normalization of Russian can be interpreted as part of the Soviet hegemonic process, which was produced and reproduced in both official institutional practice and in individual practice. The attempts by the Tatar intelligentsia to reclaim ground in the new post-Soviet social and political order can

be seen as attempts to take over the hegemonic process and the standardization of Tatar; to control the reproduction practices of ideologies and thus make *Tatar* ideologies of Tatar language and culture, rather than *Russian* ideologies of Tatar language and culture, the norm. As case studies of language ideologies and ideological debates have shown, languages are often promoted as “a crucial ingredient of national identity (and hence a central ingredient in national mobilization and nation-building)” (Blommaert 1999: 31), and Tatar in post-Soviet Tatarstan is no exception.

Language is only one marker of identity for a group, and usually part of a feature cluster that includes descent, history, culture, and religion (Blommaert and Verschueren 1998: 192); as we will see in the Tatar example, invoking one of the identity markers can usually index all the other markers in the feature cluster. Language ideologies are “thrown into high relief” by social inequalities and colonial (and thus also post-colonial) encounters (Woolard and Schieffelin 1994: 55-6), and language ideological debates are usually part of more general sociopolitical processes (Blommaert 1999: 2). When a language or language variety that indexes a subordinate group is “revalorized,” that language is usually not only a symbol of group identity, but also emblematic of “political allegiance or social, intellectual, or moral worth” (Woolard and Schieffelin 1994: 61). Post-colonial revalorization of a subordinate language can thus be seen as an attempt to manipulate what Bourdieu has called “symbolic capital” (described most fully in Bourdieu 1999). Bourdieu theorized that linguistic forms have no inherent power but rather reflect the power of the group that they index; that the legitimization of one language variety, presented as inherently better, is a misrepresentation; and that linguistic insecurity is a result of “symbolic domination.” However, case studies of linguistic performance have found flaws in Bourdieu’s theories, demonstrating that the maintenance of non-standard languages and varieties can be

interpreted according to social relations that are based on solidarity versus status, and that the heterogeneity of non-standard linguistic practice may also play a role in subordination (Gal 1989: 353-4). Tatar language reforms are meant to “add value” (or symbolic capital) to the subordinate Tatar language, and when combined with pro-Tatar language policies, the added symbolic capital can also translate into material capital: recall that bureaucrats who demonstrate a sufficient level of Tatar-language competence at one point earned more pay than their monolingual (Russian-speaking) counterparts, and that some vendors at Kazan markets are learning Tatar in order to increase their customer base to include those Tatars who prefer to transact business in Tatar.

The Tatar discourse of purity, as is common for post-colonial linguistic purism movements, is based on a “logic of oppositional identity” (Jaffe 1999: 61). Explicit linguistic ideologies can affect a language’s structure (Gal 1989: 354), and linguistic purification can be seen as one way in which subordinate languages in contact situations can become “anti-languages,” which have undergone a “Silversteinian distortion that makes a code more like itself, in this case, importantly, a self that is most distinctive from its socially dominant counterpart.” (Woolard and Schieffelin 1998: 70). In a well-known formulation of the nation as an “imagined political community,” Anderson (1987: 6-7) describes nations as constructions that are inherently limited and bounded, such that there is the set of people who are members of the nation and the set of people who are not members of the nation. Tatar identity, in particular as constructed through linguistic performance, is inextricably linked with orientation towards or away from Russian language and culture (as reified by Soviet nationalities policies), where the integrity and cultural “purity” of post-Soviet Tatars – thought by many to be necessary for the survival of the Tatar language, culture, and nation – is equated with their de-Russification. De-Russification, the removal, or “purification,” of

salient Russian influence, is expressed in various ways by Tatars in present-day Tatarstan – for example, by choices in attire, food and alcohol consumption, and sexual mores, all of which can be related to a rise in religiosity – but in this chapter my focus will be exclusively on purification and de-Russification as expressed in linguistic performance.

The Tatar ideology of linguistic purism is both explicit and implicit, and found on both the individual and collective levels. On the individual level, linguistic purism is expressed both through explicit statements that demonstrate the high level of language consciousness for Tatars in present-day Tatarstan, and through the style shifting described in Chapter One (and again in Chapter Three), where styles associated with the presentation of a conspicuously Tatar identity are also associated with a lack of language mixing, and the only Russian lexical items present in the Tatar performance style are conventional borrowings for which there is no native equivalent. On the collective level, discussions of and debates on linguistic purism are found in official language commissions, on television and the radio, and in the pages of the daily press. The majority of the examples of the Tatar discourse of purity presented in this chapter come from the post-Soviet Tatar press, in particular from daily newspapers, and are excerpted from both articles and published letters to the editor. The entire Tatarphone community takes part in this public debate on linguistic purism: articles and letters are written both by journalists and language “experts” (usually Tatar philologists) and by regular readers of the newspaper who feel compelled to express their opinions; these readers are both urban and rural, and span a wide range of ages.

The press in general plays an important role in ideological debates and in nation building. Gal (1989: 355), writing on Anderson’s work on nations and nationalism, notes that the link between language and nation is not located in the “sharing of language per se but in the unifying effects of print” and the imagined community that forms the national

consciousness is created in part by newspapers and novels in addition to speech communities. Anderson suggests that print languages lay the basis for nationalist consciousness in three ways: (1) they create “unified fields of exchange and communication,” usually somewhere between the high language used only by the educated elite and the vernacular; (2) they give language a fixed nature, which helps “build that image of antiquity so central to the subjective idea of the nation”; and (3) print languages become languages of power, such that the non-standard languages or language varieties that are not used as print languages become increasingly subordinate (1987: 44). The debates found in the daily press, which often focus reflexively on the language of the daily press, show how critical it is for “ideological and political control to be instantiated in everyday practices” (Jaffe 1999: 61). Texts generate their publics, and publics generate their texts (Blommaert 1999: 6), and in the Tatar case, this can be seen most clearly in the dialogue found in the pages of Tatar daily newspapers. The articles cited in this chapter represent just a tiny subset of the hundreds of articles published in the post-Soviet Tatar press on the subjects of language, religion, and culture in the years 1990-2001, the years covered by this literature review. The majority of pertinent articles come from several main sources: eleven years of the mainstream paper *Shäbri Kazan* ‘Kazan City’; the entire five-year run of the cultural newspaper *Mädäni Jomga* ‘Cultural Friday’; several years of the government newspaper *Watanım Tatarstan* ‘My Homeland Tatarstan’; several years of the Turkish-owned Kazan-published newspaper *Zaman* ‘Time’; and all issues of the scholarly journal *Fän häm Tel* ‘Science and Language.’ The discourse found in the articles cited here can be grouped into three overarching themes: (1) Tatar feelings of impurity because of the influence and interference from Russian language and culture; (2) purity as an explicit ideal, particularly in language; and (3) the appropriate place of Arabic loanwords and influence in “pure” post-Soviet Tatar.

Why, in the logic of language ideologies, does the purity of a language matter? In Corsica, long under French domination, language “came to stand metonymically for the totality of the struggle for nationhood” and was linked to a romantic ideology in which the language of a minority people is inextricably linked with the “soul” of that people (Blommaert 1999: 12-13). The same is true in the Tatar case, where language, culture, nationhood, and religion form an indexical feature cluster: discourse on one of these features automatically references discourse on the other features, and articles, for example, on religion can transition to discussions of language without any sort of explanatory segue, and vice versa. Fear for the assimilation of the Tatar nation – one of the most prevalent themes in post-Soviet Tatar discourse, seen in articles with titles such as *Öch mengengebe yılda tatar milläte saklanırmi?* ‘In the third millenium, will the Tatar nation be preserved?’ (Mostafin 1999) – is also expressed in discourse on linguistic, cultural, and religious assimilation. For example, we find the statement “The preservation of a language means the preservation of a people” in an article entitled *Telebezne, dinebezne saklıyk!* ‘Let’s save our language and our religion!’ (Mäjıtoı 1996); here nationhood, religion, and language are explicitly linked in a single piece of writing.

In the ideology found in the post-Soviet Tatar press, language in particular is seen as a metonymic representative of the nation and the barometer of the health of the nation, where the impurity and decline of the Tatar language are seen as representative of the impurity and decline of the Tatar nation as a whole. A typical expression of this metonymic conception claims that “...a people without a language can not be a people in the true sense – its culture, literature, customs, and rituals can only live at a time when the native language is living” (Väliyev 1992). An even more explicit expression of this linkage is found in an article entitled, quite straightforwardly, *Tel saflığı – millät pak’lege* ‘The purity of a language is

the purity of a nation' (Fäsakhov 1994). The author first draws upon a prominent Russian cultural source as authority: "The well-known Russian writer F. Dostoyevsky said, 'A language – it is the people.' So if some kind of language exists and is in use, then the nation owning that language also exists..." He then continues, "If that language, having become entangled with another language...begins to be diluted, then fear for the existence of the nation is born, and the nation steps onto the path to extinction. Therefore the purity of the mother tongue...is perhaps the basic factor deciding the fate of a nation." This is, in the logic of this discourse, due to assimilation: "A person who has lost his/her native language stops being a member of the nation...s/he has changed into a member of the nation whose language s/he knows and speaks. One or two generations after the Tatars who have forgotten their native tongue will be people who have become Russian..." Therefore, "the attempt to reanimate, cleanse and develop our native language is a battle for our nation's purity, safety, non-liquidation, and, in the end, for our independence." The logic is clear and explicit: (1) speaking or not speaking a language is the marker of a cultural and ethnic boundary, (2) if a language is a people, and the language is lost, then the people will be lost as well, and thus (3) language promotion and purification are necessary for survival of the nation.

The decline of the Tatar language is usually described in the Tatar press as externally motivated,

"the result of the evil politics of the Russian imperium to drive the Tatar nation into nothingness...Tatar has constantly been derided as a base, savage language, it was removed from all societal and governmental spheres, children's schools were purposefully Russified, and Tatar writing was forcefully moved to Cyrillic: these were some of the basic methods of the destroyers of the Tatar language. As a result of these aggressive methods against the Tatar language, our young people have stopped respecting our language, alongside the 'great' Russian language they have started to consider it weak, pitiful,

not useful, and even unnecessary. Of course, this is the result of 450 years of being colonial” (Mazhar 1996).

Due in part to these methods and factors, post-Soviet Tatar discourse, as seen in even the headlines of articles, presents the maintenance of Tatar ethnicity as difficult – *Tatar bulu bik awir* ‘It’s very difficult to be a Tatar’ (Anonymous 1999) – and a potential source of shame – *Minem uris kızı bulasım kılâ!* ‘I want to be a Russian girl!’ (Zinätullin 2000), but people are also proud – *Tatar buluim belän gorur min* ‘I’m proud to be a Tatar’ (Jälälova 1997) – and ready to do what is necessary – *Tatarlık belän gorurlanu gına jıtmi, aning mänfägat’lären jing sizganıp yaklarga da kiräk* ‘Pride in Tatariness isn’t sufficient, it’s also necessary to roll up your sleeves and get to work supporting its interests’ (Mirzayanov 1998). Recall that, as described in Chapter One, there is a derogatory name for Tatars who have assimilated to Russian culture, “Mankurts,” defined by one young Tatar as “Tatars who don’t know their own language or history, their own traditions, who don’t understand or respect their ancestors” (personal communication).

The definition of pure Tatar is a negative one, and can be interpreted as based on Jaffe’s “logic of oppositional identity”: pure Tatar is, quite simply, Tatar that is not Russian and expresses no salient Russian influence.² Doctrines of linguistic purism often “close off non-native sources of innovation, but usually selectively, targeting only languages construed as threats” (Woolard and Schieffelin 1994: 64). This is true in the Tatar case as well, where the language of the culture and government perceived as a threat to the integrity of the Tatar nation is the target of purification, both in language reform and in the discourse of purity,

² I say “salient” here because Tatar speakers and adherents of this discourse of purity do not articulate an awareness of all kinds of Russian interference: as will be seen in the remainder of this chapter, the focus of purist discussions is usually on what I like to call “sounds and nouns” – phonetic interference and borrowed nouns. Other kinds of interference, for example, calquing or the code-mixing or borrowing of words in other grammatical classes, receive little to no attention.

while influences from languages and cultures not perceived as threatening are either interpreted as congruent with the post-Soviet Tatar identity or not salient to purists.

I will illustrate this with two anecdotal examples, both from individual rather than public discourse. The first is a comment made to me by a young professional in his mid-30s, a frequent visitor to the Tatar Social Club who seems dedicated to speaking as little Russian as possible – in all the months I knew him, I never heard him say one Russian word, and he once reprimanded me for speaking Russian with a young Tatar woman who works in the library, saying to me, “She’s a Tatar girl, with a good Tatar name, you should speak Tatar with her, and not Russian.”³ He brought up the concept of pure Tatar with me when we were discussing a Finnish Tatar who was visiting Kazan for a short time – this Finnish Tatar’s home language growing up in Helsinki was Tatar, the Mishar dialect (albeit transplanted to Finland), and he spoke only a few words of Russian, even after months in Kazan. My friend remarked: “Do you know why I love to hear [him] speak? Because there is not even a hint of Russian, he uses only good Tatar words, the words we all used before the Revolution. Hearing him talk, it is like music to my ears.” This comment is of particular interest in two ways. First, the Mishar dialect, at least as spoken within the boundaries of the Russian Federation, is considered extremely low-prestige and somehow lacking the “richness” of the Middle Volga/Kazan Tatar dialect; this man’s lack of Russian interference, perhaps in combination with the prestige associated with Finland (which is associated with things European), seemed to supersede the stigma associated with the Mishar dialect.

³ The woman in question is Russian-dominant and will not speak Tatar with patrons, although it is clear that she understands them. When addressed in Tatar, she will respond in Russian, and most patrons will accommodate and switch to Russian with her; the only exceptions I ever saw were a few people over 60 and this particular club member.

Second, due to the fact that the Helsinki Tatar population emigrated from Russia approximately 100 years ago, before Soviet-era language engineering, this Finnish Tatar does not use most of the Russian borrowings that are standard in present-day Tatar (at least, those later Russian borrowings that were not adapted to Tatar phonotactics), but rather their pre-Revolutionary counterparts, most of which are in fact Arabic and Persian loanwords. Even so, his speech was assessed as a model of pure Tatar, our first indication of the selective nature of the Tatar discourse of purity – Russian words, borrowed from a language “construed as a threat,” have no place in pure Tatar, but words borrowed from Arabic and Persian, languages and cultures that were historically highly influential for centuries, do.

The second example also involves a decidedly pro-Tatar young male, a 30-year historian and increasingly observant Muslim who studied Arabic in Eastern Africa for several years and even changed his name because his original name, although common among Tatars since the 1930s, is borrowed from English and thus wasn't sufficiently Tatar for him – his new name is both traditionally Tatar and explicitly religious. After hearing me interviewed on a Kazan radio program, he requested that our mutual friend introduce us, and when we met, he spent several minutes telling me about what his impressions had been while hearing me speak Tatar on the radio. For him, the most prominent aspect of my Tatar was its lack of Russian interference or influence: I had apparently sounded to him like a “young Tatar village girl” who had been “protected” from both the Russian language and from Russians, as if I had lived a “clean village life” and had never spoken a word of [presumably unclean] Russian. Several people with whom I became acquainted after my radio interviews told me that they had interpreted my accent as native, albeit a speaker of a

different dialect,⁴ but this man was the only one to explicitly relate my lack of Russian accent to impressions of pure Tatar girlhood. This idealization of village life, and of the village as protected enclave of Tatariness, is quite a common one, particularly noticeable in Tatar-language comedies and musicals produced in Kazan theaters (Faller 2001). After learning that I was studying and recording Tatar, person after person would invite me to visit their village⁵ so I would be able to hear pure Tatar as spoken in all-Tatar villages where Tatar is spoken all day, every day, far away from the “corrupt” city and “street language.”⁶ Even my Russian cab driver one day suggested to me that if I really wanted to learn Tatar, I should head to the countryside: “You’re not going to hear any real Tatar here,” he said.

2.2 “Impure” Tatar

Before examining the purification movements associated with Tatar linguistic purism, I will discuss some of the “impure” Tatar that Kazan’s Tatarphones are surrounded by daily, since this “tainted” Tatar is the backdrop against which “pure” Tatar is set. Although Tatar has been elevated to the level of governmental language, officially on a par with Russian, and is correspondingly seen and heard more in public spaces than in Soviet times, there do not appear to be a sufficient number of fully proficient Tatar speakers to

⁴ I should note that it is not that my Tatar is, or was at the time, anything like that of a fully competent native speaker. While my accent is good, my intonation patterns are wrong, my syntax is limited, and my speech not without grammatical errors. People’s willingness to perceive me as a native Tatar says less about my fluency, and more about both their lack of exposure to any accent in Tatar other than Russian, and about the Tatar-language (in)competence of many young Tatars. See Woolard 1989: 70ff for interesting parallels in Catalonia.

⁵ If they were city-born, the invitation would be to the village where their closest relatives live.

⁶ In point of fact, young people in villages do not speak only Tatar all day every day, as became immediately apparent upon my visits.

translate or produce all of the public Tatar that is meant to serve as the equivalent to Russian. What this means is that, at least in Kazan, Tatar that is poorly translated from Russian, misspelled, and grammatically incorrect according to the literary standard is a constant presence. One example is programs at the opera, ballet, and symphony: more than once, I have seen Tatars turn away from the Tatar half of a bilingual program to the Russian half, and complain that not only was it clear that the Tatar text was a translation from Russian and not originally composed in Tatar, but that the translator had done a poor job of translating. However, they are often unable to articulate precisely what the grammatical incursions are when they are outside the scope of phonetic interference or semantic mismatches in translation. Letters to the editor complain that both street signs and signs for bus and tram stops are incorrectly spelled or translated, as is signage at museums and galleries. For example, my first week in Tatarstan, I visited an art and culture museum in downtown Kazan. As I stood looking at 19th century leather boots, the man next to me began agitatedly complaining (in Russian) to the museum guard about all the grammatical mistakes in the Tatar museum signage. She told him that he should be thankful that there were exhibit signs in Tatar at all, that only a few years previously everything had been in Russian, and that while there were probably many mistakes, it was better than having no Tatar at all. There are Tatar language teachers at the elementary and secondary level who are unable to speak grammatically: not all of them, by any stretch, but I personally encountered three teachers who made very basic mistakes when speaking Tatar, including the use of the absolutive case for pronouns with postpositions that govern nouns in the absolutive case but require pronouns in the genitive case (e.g., *min öchen* ‘for me’ in lieu of *minem öchen*). Additionally, there are university lecturers and professionals in various fields who do not have a fully developed formal register, or professional or educational vocabulary (this is due

to a lack of developed professional lexicons and specialized education in Tatar). There are stories, perhaps apocryphal, of Tatar philologists at the university who are forced to consult with village-born secretaries on questions of vocabulary and syntax, but more easily observed is the inability to speak about one's profession or field of specialization using Tatar only, causing either a reversion to Russian, or grammatically incorrect and halting Tatar if the situation does not allow Russian to be used. In terms of cultural production, there are Tatar musicians who compose songs without knowing Tatar, or who sing in Tatar with Russian accents: one article, *Urıs telle, Tatar küngelle Oskar* 'Russian-language, Tatar-souled Oskar' (Jälilova 2000) profiles (mostly positively) a young half-Russian, half-Tatar composer who does not know Tatar but has nonetheless composed melodies for over one hundred Tatar pop songs. When asked about his lack of Tatar knowledge, he replies, "I don't need a text's word-for-word translation to write a song. If its author gives me a few words about its character and mood, that is sufficient." Other actors in the field of cultural production can also be seen as deficient in their Tatar-language skills: a recent report tells of a new dictionary currently under preparation at the Tatarstan Academy of Sciences' Language, Literature and Art Institute that is meant to "improve the spelling and speech of Tatar journalists," apparently by means of focusing on the correct usage of polysemous words and homophones (RFE/RL September 6, 2002).

2.2.1 Phonetic interference

Tatar has nine phones that are not found in Russian, all of which are phonemic, and these phones can cause problems both for Russian speakers and for Tatars who have heavy Russian interference. Table 1 shows these Tatar-specific phones, their orthographic equivalents in the Tatar Cyrillic alphabet, and their closest Russian equivalents used as standard substitutes:

Table 1. Tatar-specific phones

<i>Tatar-specific phone</i>	<i>Cyrillic letter</i>	<i>Closest Russian equiv.</i>	<i>Cyrillic letter</i>
[æ]	‘ə’	[a]	‘a’
[ø]	‘ə’	[o], [e]	‘y’, ‘e’
[y]	‘Ÿ’	[u]	‘y’
[q]	‘к’, ‘кь’	[k]	‘к’
[ɣ]	‘г’, ‘гь’	[g], Ø	‘г’, Ø
[dʒ]	‘ж’	[ʒ], [z]	‘ж’, ‘з’
[h]	‘h’	[x]	‘x’
[w]	‘в’	[v]	‘в’
[ŋ]	‘h’	[n]	‘h’

The above Tatar-specific vowels are all front vowels that are necessary for the correct realization of vowel harmony: for example, [æ] is found in two case-marking suffixes, locative *-DA* and ablative *-DAŋ*, and in the plural suffix *-LAR*.⁷

Some kinds of Russian phonetic interference are more salient than other kinds: the phone [h] in particular seems to be especially emblematic of a sound subject to phonetic interference, and entire articles are devoted to its mispronunciation as Russian [x]: *А*

⁷ These suffixes are subject to both vowel harmony and consonant harmony: as is traditional in descriptions of Turkic languages, the vowels and consonants that assimilate harmonically are written here in capital letters.

tatarnıng rünjetelgän mesken “h” awazı ‘O the offended, poor sound “h” of Tatar’ (Gallämov 1998); *Ber awaz tiräsändä* ‘About one sound’ (Kärimi 1999); *“H” awazına hälakät yanıy* ‘The letter ‘h’ burns and perishes’ (Mazhar 1996). An editorial in the monthly language supplement *I Tugan Tel* ‘O Native Language’ that appears in the weekly cultural newspaper *Mädäni Jomga* rails at the general mispronunciation of Tatar names that have Tatar-specific sounds, particularly family names that are now used as street names and tram stops in Kazan: for example, [x] for [h], [u] for [ø], and [a] for [æ] (Anonymous 2000). And the author of an article entitled *Telebezıng maturlıgın saklıyk!* ‘Let’s save the beauty of our language!’ complains that the the voiced fricative [ɣ] is disappearing at the beginning of names such as Gabdulla, making them less Tatar (Kärimullina 2001).

2.2.2 Spelling and translation errors

Spelling problems are another aspect of “impure” Tatar that are made explicit in public discourse in newspaper articles and letters to the editor. In an article entitled *Tatarcha soyläshäbezme? Dörsen yazabızmi?* ‘Are we speaking Tatar? Are we writing correctly?’ a prominent scholar describes spelling mistakes made by academics writing in one of the leading newspapers in the city (Äydi 1999). Khujamät (1999) tells of a prominently displayed billboard in downtown Kazan advertising an international company: the billboard says “good work” in several European languages, one of them Tatar. However, while the phrase is spelled correctly in all the other languages, in Tatar there are two spelling errors – there are six letters in total, and two of them are wrong. And in a letter to the editor given the title *Tatarcha döres yazıyk* ‘Let’s write Tatar correctly’, one Kazan resident writes, “I walk around and see lots of mistakes in how Tatar is written in store and street names. Russian words are

written correctly, but why in Kazan, the capital of Tatarstan, are Tatar words written incorrectly?” She details several obvious mistakes found in her neighborhood, and concludes, “Someone who can’t write names in two languages maybe needs to research his/her correct spelling” (Fatiykhova 1998).

2.2.3 Code-switching

Tatar linguistic purism is opposed to the use of Russian words in Tatar, both as conventional borrowings in Tatar and as code-switched text; conventional Russian borrowings will be discussed in 2.3ff. Bilingual Tatars usually have a range of speaking styles that can be seen as a continuum with pure Tatar on one end and pure Russian on the other end, with styles where there are various degrees of language mixing found in between (cf. Figure 2 in section 1.4 of Chapter One). In Chapter One, I demonstrated that Tatar style shifting can be interpreted according to Bell’s (1984) theory of audience design, such that each style is designed for the appropriate audience and situation. For example, “Tatar on-stage” style is appropriate for public speaking at Tatar cultural events; pure Russian is appropriate for speaking with Russian monolinguals, and some sort of code-switched style is appropriate for certain situations when one is relaxing with friends (although in other situations “Tatar-preferred” style is more appropriate). This language mixing is by no means limited to younger Tatars, but it is often perceived as an aspect of youth speech, leading some to believe that the code-switching speakers are incapable of speaking in Tatar without recourse to Russian, which may or may not be true, and it is highly stigmatized by many. For example, in an article entitled *Ike tellek, imesh* ‘Bilingualism, kind of’, the author tells of an incident where he sits on a train near two older women who are code-switching for the entirety of the trip – he finds their language “repulsive,” and feels compelled to reprimand them. He tells

them, “You, old women...you are returning to the Tatar territory of Archa [Russian *Арск*], the soil of Gabdulla Tukay, and you are muddying Tukay’s language, crushing it” (Äkhmatjanov 1997). Tukay is sometimes known as the Tatar Pushkin, and is the most beloved national poet; his most famous poem, *I Tugan Tel* ‘O Native Language’, has been set to music and is now the unofficial Tatar anthem. In this discourse, the code-switching of Russian in with Tatar is seen as making the language impure, and this impure language, in turn, “muddies” the pure soil of the Tatar territory where Tukay, the ultimate symbol of Tatar culture, once walked. This relationship of pure language and pure rural space echoes the young Tatar who connected my Tatar speech, which is without Russian phonetic interference, to the “clean” village life of a “pure” Tatar girl: unsoiled language, unsoiled girlhood, unsoiled soil – these are the ideals of the purist doctrines, and, among others, the goals of Tatar linguistic purification and de-Russification.

2.3 Tatar purification movements

In addition to linguistic purification at the individual level, which takes its form in a style shifting continuum with a de-Russified “pure” Tatar performance style at the extreme Tatar end, there are several purification movements in present-day Tatarstan. In section 2.3.1, I will describe the current orthographic reform and the symbolism of the various Tatar alphabets used today, and in section 2.3.2, I will describe the purification and de-Russification of the Tatar lexicon. In addition to these two means of purification, another suggested procedure for de-Russification is the re-Tatarization of conventional Russian borrowings. For example, one philologist recommends that the Tatar suffixes added to Russian words should follow the rules of Tatar vowel harmony, and that conventional

Russian borrowings in Tatar should be spelled phonetically, e.g., тавар (*tavar*) rather than товар (*tovar*) for [tavar] ‘commodity.’ These Russian borrowings, he says, should also follow the rules of the native Tatar phonological system and should be “Tatarized” if they do not, e.g., words that begin with unacceptable consonant clusters should have vowels placed at the word’s beginning, such that the word *sbtani* ‘trousers’ would then be realized in Tatar as *ışbtani*⁸ (Äkhmät’yanov 1999). This initial-vowel insertion would be parallel to the word-initial vowel insertion that often takes place in colloquial Turkish with borrowings that violate native phonology by beginning with certain phones, usually nasals or liquids, e.g., *limon* > *ilimon* ‘lemon.’ Some of these pronunciations with word-initial epenthetic vowels are now standard, e.g., *orospu* ‘harlot’ < Persian *ruuspü* (Lewis 1968).

While it is unlikely that most conventional Russian borrowings will start to be spelled phonetically, or altered to fit native phonological rules, other “Tatarizing” modifications do take place, most particularly in the media. Newspaper and journal editors play a real and important role in the purification and de-Russification of texts, often without authors’ knowledge or permission. One author states, “...let’s take the word *Rosıya* [‘Russia’]. I use it in my writings without alteration. But in the editorial process, in the preparation to print my article, they change it to *Rüsiyä, Räsäy*”⁹ (Minhaj 1997). Editors also sometimes replace conventional Russian borrowings with equivalent now-archaic Arabic borrowings from pre-Revolutionary Tatar. For example, one journalist had the term *khäyer-khalik* ‘benevolence, goodwill’ inserted in one of her articles: when I asked her what the word meant, as I was not familiar with it, she admitted, with some embarrassment, that she herself did not know. It

⁸ Due to hundreds of years of extensive contact with and borrowing from Russian, these native phonological rules have been violated for so long that they are probably no longer perceptible to the average Tatar speaker, particularly one who is bilingual in Russian, which is to say, the vast majority of Tatar speakers.

⁹ These pronunciations fit the traditional Tatar phonetic and phonological systems.

has been suggested that this editorial process is one of the primary means of the reintroduction of archaic Arabic and Persian borrowings: “It would not be a mistake to say that the phenomenon of the return of Arabic and Persian loanwords to our language began first of all in the pages of the daily press” (Safiullina and Fyodorova 2000). The return of Arabic and Persian loanwords is not entirely uncontroversial, and will be discussed in detail in section 2.3.2 below.

There are several factors that allow the inclusion of Arabic and Persian influences as part of the post-Soviet Tatar identity, one of which is the relationship between language and religion. Recall that it is not unusual to find romantic ideologies that link language and the “soul” of a people. In this ideology, linguistic purity is also related to purity of the soul, and therefore linked with religion (and in this manner also linked with the Arabic language and culture that are associated with Islam). Officially banned and suppressed during Soviet times, and persecuted by Russians for centuries before that, Islam is experiencing a religious revival in Tatarstan, and the number of active mosques in Tatarstan has increased dramatically: from 19 in 1989 to 750 today. The number of Tatars identifying themselves as practicing Muslims has increased as well, to 67% of urban Tatars and 86% of rural Tatars (Zolotov 1999). Along with more open religious worship has come more open discussion of religious issues, and many of the articles that discuss religion link it explicitly with language, for example: *Din belän tel bezne saklap kalgan* ‘Religion and language preserved us’ (Mökhbullin 1995), and *Telebezne, dinebezne sakliyk!* ‘Let’s save our language and our religion!’ (Mäjitoŭ 1996). When language is not explicitly linked with organized religion, then it is often related to the “soul” or spiritual issues, e.g., *Anam tele – janim tele* ‘My mother tongue is the language of my soul’ (Nizamov 2001). In this logic, it is important to use words correctly, “because a word cultivates the soul... ‘Ruining the language ruins the soul,’ say the people” (Shäfi 1992).

Work on cultural and linguistic preservation is seen as holy – for example, one article discussing the publication of a book on Tatar national clothing (related only most distantly to language and religion) is entitled *Telebez öchen dä, mädäniyätebez öchen dä izge esh* ‘Holy work on behalf of our language and our culture’ (Ramazonva 1998). A student at Kazan State University, in the middle of an article explaining his decision to study in Tatar, writes, “We know that the Tatar language, together with the Islamic faith and the Tatar people, has undergone much difficulty...The future is in our hands. The growth and enrichment of the Tatar language is our holy debt...” (Sadriyev 1997). Another KSU student, writing several years later on the same topic, touches on the same theme: “A native language is...the means of expression the most fine and holy feelings in a person’s soul...Keeping away from the native language and its foundation deprives the young people who purposely do not learn the native language of national awareness, personal pride, and of the ethical and moral sources of our people” (Khäyrullina 1999).

Religion is accessed through the Tatar language – while the Koran is written in Arabic only (direct translations are expressly forbidden), summaries of the Koran and other religious teachings, including sermons by imams and classes in *mädräsäs* are in Tatar. Speaking Tatar properly is then seen as necessary to properly access the spiritual “purity” to be found in these teachings. “The Islamic faith has been given praise and glory. It is considered to be the most pure, the most humane of faiths by members of other religions” (Shäykiyev 1993). It is this inter-relatedness of language, nation, and religion, along with the historical significance of Arabic and Persian as languages of the world-wide Islamic culture to which the pre-Revolutionary Tatars belonged, that informs both orthographic and lexical reform in present-day Tatarstan.

2.3.1 Orthographic reform and symbolic alphabets

The purification movement that is currently most conspicuous and politically topical is Tatarstan's recently begun shift from the current Cyrillic alphabet (which includes six Tatar-specific letters, shown above in Table 1) to a Latin-based alphabet, the fourth alphabet in a



Figure 5. *The Latin-based Alphabet.*

period of less than 100 years. The new alphabet was ratified in September of 1999, and the ten-year transition period began with street signs and school programs in the fall of 2000. Figure 5 shows the cover of a textbook for schoolchildren published in 2000, designed to teach literacy in the new alphabet (Wagizov and Wälitova 2000), and Figures 6 and 7 show the alphabet as presented in this textbook. Orthographic battles

are common in situations “where identity and nationhood are under negotiation”; this is because “orthographic systems cannot be conceptualized simply as reducing speech to writing, but rather...are symbols that carry historical, cultural, and politicized meanings” (Woolard and Schieffelin 1994: 65). Statements about language, along with linguistic reforms, serve as signifiers for social and political realities; they are metonymic (Gal 1998), and seemingly innocuous language reforms actually often serve as symbolic gestures with deeper, sometimes provocative, meanings. For example, the American Ebonics debate of 1996-1997

Басма хәреф	Язма хәреф	Әйтелеше	
		латиницада	кириллицада
A, a	<i>A a</i>	a	а
Ә, ә	<i>Ә ә</i>	ә	ә
B, b	<i>B b</i>	be	бэ
C, c	<i>C c</i>	ce	жэ
Ç, ç	<i>Ç ç</i>	çe	чэ
D, d	<i>D d</i>	de	дэ
E, e	<i>E e</i>	e	е (э)
F, f	<i>F f</i>	ef	эф
G, g	<i>G g</i>	ge	гэ
Ğ, ğ	<i>Ğ ğ</i>	ğı	гъ
H, h	<i>H h</i>	he	һэ
I, ı	<i>I ı</i>	ı	ы
İ, i	<i>İ i</i>	i	и
J, j	<i>J j</i>	je	жэ
K, k	<i>K k</i>	ke	кэ
Q, q	<i>Q q</i>	qı	къ
L, l	<i>L l</i>	el	эл

Figure 6. Tatar alphabet a-l from Latin Grafikası Nigezendä Älifba.

Басма хәреф	Язма хәреф	Әйтелеше	
		латиницада	кириллицада
M, m	<i>M m</i>	em	эм
N, n	<i>N n</i>	en	эн
N̄, n̄	<i>N̄ n̄</i>	eñ	эң
O, o	<i>O o</i>	o	о
Ө, ө	<i>Ө ө</i>	ө	ө
P, p	<i>P p</i>	pe	пэ
R, r	<i>R r</i>	er	эр
S, s	<i>S s</i>	es	эс
Ş, ş	<i>Ş ş</i>	şa	ша
T, t	<i>T t</i>	te	тэ
U, u	<i>U u</i>	u	у
Ü, ü	<i>Ü ü</i>	ü	ү
V, v	<i>V v</i>	ve	вэ
W, w	<i>W w</i>	we	вэ (у)
X, x	<i>X x</i>	xa	ха
Y, y	<i>Y y</i>	qısqa i	кыска и
Z, z	<i>Z z</i>	ze	зэ

Figure 7. Tatar alphabet m-z from Latin Grafikası Nigezendä Älifba.

was “about a proposal to revalue the symbolic status of a nonstandard variety, and the fury of the response to that proposal must be seen as having roots in a broader reaction” (Collins 1999: 223). And indeed, the Tatar orthographic reform, which is also a proposal to revalue the symbolic status of a minority language variety, is an increasingly sensitive topic, especially with political and legislative opposition coming from government institutions and legislative bodies in Moscow since the fall of 2001. The legal status of the transition to the new alphabet changed frequently during the first half of 2002, and is now being reported not only in Tatarstan’s press but also nationwide in Russia. Most recently, on November 15, 2002, the State Duma passed the third and final reading of an amendment that prohibits the use of non-Cyrillic scripts by state languages in the Russian Federation – this amendment was provoked directly by the Tatar alphabet shift. Reactions and editorials in the non-Tatar press can be inflammatory and predictive of certain doom; an early editorial, much quoted in the Western press digests of the time, stated, “Every second resident of Tatarstan is a Russian. And at the same time, two thirds of Russian Tatars live outside Tatarstan. In Moscow alone live 600,000. Did they think about these people...when preparing this ‘reform’? Inter-ethnic conflicts often begin with language. And they end with a lot of blood. Are they really preparing this fate for us?” (Alimov and Iusin 1999). There is a general awareness of the symbolic nature and political sensitivity of the orthographic reform: a significant percentage of young people I interviewed opted to skip all questions relating to the transition to the new alphabet.

The official justification for the new alphabet, as found in the Tatar version of the law ratifying it, is that it “more accurately represents the particular characteristics of the Tatar language,” and will allow “entry into the system of world communication” (Shäymiyev 1999). Some supporters of the transition to the new alphabet explicitly refer to language

purity as part of their justifications: for example, one letter to the editor – given the title *Latin älifbası telne chıstartırmi?* ‘Will the Latin alphabet clean the language?’ – ends this way: “Unfortunately, over the years, the Russian alphabet as used has done great injury, without end, to the purity of the Tatar language. Because of the lack in this alphabet of four letters that are necessary to the Tatar language, our language continues to degenerate from year to year” (Galiyev 1998). To be frank, it is not clear to which four letters Galiyev is referring: with the additional Tatar-specific six letters and the use of hard signs to denote post-velar consonants, all of the phonemes of the Tatar language except for /w/ have distinct orthographic representations. Also unclear is how the new Latin-based alphabet will aid Tatar’s entry into worldwide communication, more specifically, the internet, one of the stated aims of the change. The new Latin-based alphabet (cf. Figures 6 and 7) has eight letters not found in English, Spanish, French or German, and five letters not found in Turkish. Therefore, new computer fonts will need to be devised for the new alphabet, and users of the new alphabet will not be able to easily interact using existing computer programs until those fonts are widespread. Although both supporters and opposition alike perceive the new alphabet as a statement of pan-Turkic solidarity, due in great part to the new alphabet’s resemblance to the standard Turkish alphabet, and as a move toward the West, based on its resemblance to the Latin-based alphabet of many European languages, it is important to note that the new alphabet is still Tatar-specific. Most notable are the letters chosen to represent Tatar’s front vowels. The high unrounded vowel pair is represented by an ‘ı’ with a dot for the front version, and an ‘i’ without a dot for the back version; this is identical to Turkish orthography. However, while [u] and [y] are represented by ‘u’ and ‘ü’ respectively (also identical to Turkish), the other front/back pairs are not distinguished by diacritics but rather with separate letters, letters retained from the Tatar-specific Cyrillic alphabet. So

instead of ‘a’/‘ä’, we find ‘a’/‘ə’, and instead of ‘o’/‘ö’, we find ‘o’/‘ə’. The chosen orthography can thus be read as a metonymic representative of political and cultural orientation: it states that while the Tatars are aligning themselves with Turkey and the West, they are still distinct from them; as a culture, as a political entity, as a nation. Yurchak (2000: 414), writing on the renaming of newly private businesses in 1990s Russia, notes that the process of renaming public space introduces a “particular new version of social reality” where the authors of the renaming “strive to impose themselves as the legitimate authors, owners, and masters of this reality.” In this way, the post-Soviet practice of renaming can be seen as parallel to the Tatar post-Soviet orthographic reform, where the new alphabet can be seen as not only reflecting a new social reality, where Tatars have more political control and greater sovereignty than in the Soviet era, but also as constructing that new social reality; it is a “linguistic innovation designed not only to reflect but also to initiate changes in the social world” (Yurchak 2000: 420).

Some supporters of the orthographic reform are explicit about its symbolic nature, and about its cultural and political components. Writing several years before the new alphabet was legally ratified, one author noted that “the majority of Turkic-speaking peoples, taking Turkey as an example, long ago started moving to a Latin alphabet. This state has helped them move closer to the cultural richness of Europe, and to its political possibilities” (Mäjitov 1996). Even more explicit are the statements in an article entitled *Möstäkiyl däülätkä üz khäräfläre* ‘For an independent government, its own alphabet’; the author argues that the shift to a Latin alphabet will facilitate the schoolchildren’s study of the languages of Europe and America (thus implying that knowledge of these languages is desirable), and goes on to state, “[The] Latin [alphabet] will strengthen our government’s independence...For a sovereign government, a “sovereign” alphabet too is necessary” (Mirsäyetov 1997).

The orthographic reform, like the post-Soviet practice of renaming of businesses, uses two semiotic functions to produce “complex cultural meaning”: metaphor and metonymy (Yurchak 2000: 416). The new alphabet provides a metaphorical link to both Western languages and to Turkish by means of shared symbols (here, letters). This metaphorical link then becomes metonymic, and the new alphabet is thus “connected to a whole variety of ideas, identities, commodities and lifestyles” that are associated with the languages and societies with which it is metaphorically linked (ibid. 416). These semiotic functions are used not only with the new Latin-based alphabet, but are also demonstrated by both the use of and interest in other alphabets historically used by Tatars (and their ancestors). The new Latin-based alphabet is the fifth Tatar writing system in recorded history: in pre-Islamic times (approximately 5th century C.E. onward), there was a pan-Turkic runic alphabet, which was superseded by a modified Arabic script that was brought into use soon after the Bulgars’ 10th century conversion to Islam. For centuries, Tatar was written in a series of modified Arabic alphabets, and those Tatars who were educated enough to read and write had an Islamic religious education, and therefore also read and wrote Classical Arabic. In 1928, Tatar, along with the Turkic languages in Central Asian Soviet republics, was shifted to a modified Latin alphabet; in the Soviet Union of this time, the Latin alphabet was considered to be the “alphabet of revolution” (Kreindler 1995). The change in orthography from Arabic to pan-Soviet Latin was one way to try to reorient the Soviet nationalities away from outside influences, in particular Muslims in the rest of the world, who were still using Arabic orthography (Grenoble and Whaley 1998; Kirkwood 1989). After ten years of this Latin-based alphabet, known as in Tatar as *Yangalif*, in 1938 Tatar was shifted to the modified Cyrillic alphabet described above as part of the Soviet Russification

process, with the idea, in part, that the shift would facilitate literacy and bilingualism in Russian.

The interest in and use of these various historical alphabets in present-day Tatarstan can be linked metaphorically and metonymically to an interest in the cultures, politics, and social realities associated with these alphabets. For example, research on and use of the runic alphabet is linked both with pan-Volga and pan-Turkic separatist sentiment and with theories of Tatar ethnogenesis that have either arisen or resurfaced in the post-Soviet era (ethnogenesis theories were discussed in section 1.1 of Chapter One). These theories, found

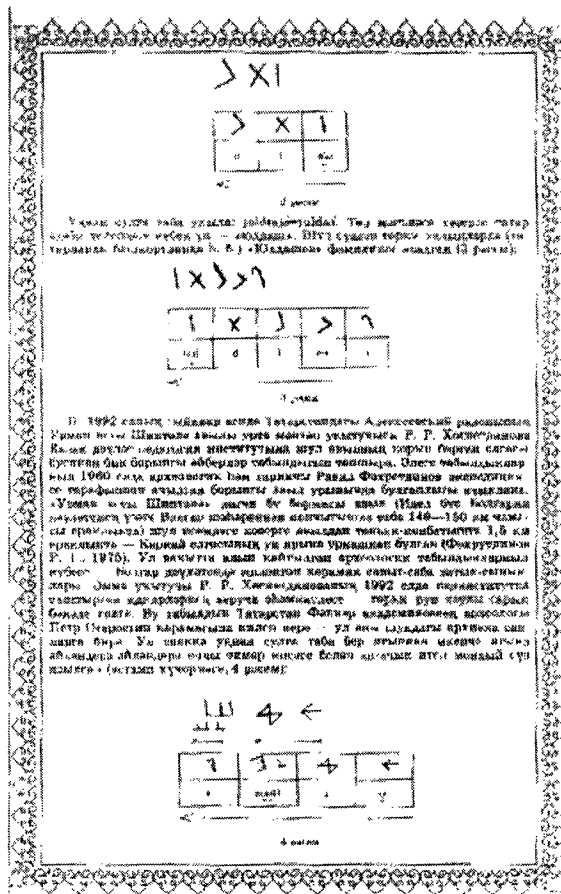


Figure 8. Excerpt from an article on runic writing (Kurbatov 2000).

in tracts such as *My – bolgary, a ne tatory* ‘We are Bulgars, and not Tatars’ (Khalil 2000) and *Tatar-Törki Etnogezı* ‘Tatar-Turkic Ethnogenesis’ (Zakiev 2000), written by a prominent Tatar linguist, are clearly attempts to use history to legitimate the Tatar presence in the Middle Volga region. Although these theories of ethnogenesis posit different origins for the Tatar people, what they have in common is both the distancing of the Tatars from the Mongolian “Golden Horde” and attempts to show that Tatar habitation of the lands that are now part of the Russian

Federation predates that of Russians. Runes, and fonts that resemble runic writing, are used to show the antiquity of the Tatar nation, and antiquity is seen as necessary to the nationalist

project (Anderson 1987). Figure 8 is a page of an article from the journal *Miras* ‘Heritage’ that is devoted to runic writing and the deciphering of ancient monuments (Kurbatov 2000). Figure 9, also taken from the journal *Miras*, shows the symbolic connection between runic writing and Tatar history: the figure is the title page that introduces the section of the journal devoted to Tatar history, and the rubric title, *Borıngı babalarıbız tarikhı* ‘The history of our ancient grandfathers’, is written in a Cyrillic Tatar font that is designed to resemble runic writing.



Figure 9. Rune-style Tatar font (*Miras* 2000).

Even when runic writing is incomprehensible as orthography to present-day Tatar speakers, it is immediately recognizable as a symbol of the *Idel-Ural* ‘Volga-Ural’ pan-Turkic separatist movement. For example, in the spring of 2001 one club member, long-associated with the Idel-Ural movement, designed a fund-raising t-shirt to be sold to nationalistically oriented Tatar youth; the slogans on the shirt were written in both the Tatar Cyrillic alphabet and in the runic orthography.¹⁰

¹⁰ Unfortunately, I was only allowed a glimpse of this t-shirt as it was en route to the store where it was going to be reproduced, and was told explicitly by its designer that, even though he liked me as a person, I could not buy a shirt, as it was “for Tatars, and Tatars only.”

Arabic orthography, banned in the earlier part of the Soviet era (during which time Tatar books written in the modified Arabic script would be confiscated and burned), is also used for its metaphorical and metonymic associations. Most Tatars are illiterate in this script, sometimes known as *iske tatar* ‘old Tatar’; even so, there are two newspapers based in Kazan

that are printed in this Arabic script, *Iske Imlyä* ‘Old Orthography’, which has a masthead partly written in Cyrillic that explains that it is a *Garäp grafikasında tatar telendü gazeta* ‘Tatar-language newspaper in an Arabic-based alphabet’, and *Iman* ‘Belief’, a religious newspaper that claims to have a circulation of 10,000.¹¹ Figure 10 shows



Figure 10. The religious newspaper *Iman* ‘Belief.’

the top half of the front page of this newspaper, which is hand-calligraphed in its entirety (4 pages), while *Iske Imlyä* is written in a computerized Arabic font with added Tatar-specific diacritics.

Like those young Tatars described in Chapter One who, in the early days of my fieldwork, refused to accommodate my lack of Tatar-language skills by speaking Russian, the

¹¹ This number is extremely suspicious; *Tatar Ile*, by comparison, a reasonably popular paper and one that is published in an alphabet that is comprehensible to modern Tatars, only claims to have a circulation of 2000.

authors and publishers of these newspapers are also clearly choosing ideology over communication. Although religious education and literacy in Arabic is increasing in Tatarstan, knowledge of the classical Arabic script does not automatically produce literacy in the *iske tatar* Arabic script, which has both Tatar-specific diacritics and orthographic ambiguities. I personally am aware of only one person who is capable of easily reading texts in this script, and she is a philologist who engages in historical research. The question then arises: who is the intended audience for these newspapers? The currently available readership is tiny, at best; therefore, a more logical conclusion is that the existence of a newspaper in Arabic script, published in distinction to the Cyrillic script that is nearly identical to that used for Russian, is more important than any of the information that is contained within the newspaper. It is an example of construction of “oppositional identity,” here linked to the opposition of feature clusters: Tatar-Muslim is contrasted with Russian-Christian/atheist. Here once again we see the selective targeting of linguistic purism, where non-native Russian is being rejected, but non-native Arabic influence embraced. This Arabic and Islamic influence is also embraced by the Tatar Social Club, which during the 2000-2001 season hosted several talks by religious speakers as well as an “Arabic friendship evening.” The club’s interest in Middle Eastern countries and cultures is explicitly expressed in its name (recall that “Tatar Social Club” is a pseudonym that I have chosen in order to protect the identity of club members), and the club logo is written in Arabic orthography. One club organizer in particular, who is also involved in the Kazan religious community, likes to make sure that this logo, printed on a piece of paper that is affixed to the wall during meetings, is visible in all group portraits taken at the club.

Arabic script as a symbol of Tatariness is found not only in the linguistic marketplace, but also the actual marketplace. Figure 11 shows a bottle of a type of liqueur called *bälzäm*

manufactured by the *Chistopol'skii Likero-Vodochnyi Zavod*; the brand name of the liqueur, *Chistoi* 'clean', is written in a Cyrillic font that is designed to look like Arabic. The name *Chistoi* is Russian, not Tatar, but the label is clearly signifying that the brand is Tatar and not Russian, and thus a brand to be purchased by those Tatars who prefer to buy Tatar products. The bottle was presented to me by Gälimä, the woman whose style shifting was described in section 1.4 of Chapter One, and was part of a parcel of symbolically Tatar food products that were given to me at the train station when I was departing from a follow-up



Figure 11. Chistoi (Чистой) *bälzäm*

fieldwork visit in the spring of 2002: other items included *chäk-chäk*, known as the “Tatar national dessert,” and two kinds of tea packaged by a company called *Teastan*, a play on words that invokes both ‘tea’ (which is not *tea* but *chäy* in Tatar and *chay* in Russian) and ‘Tatarstan.’ Gälimä was well aware of the symbolic nature of the gifts, and even apologized that one of the kinds of tea was more Bashkir-style than Tatar-style. Many politically and culturally

involved Tatars are aware of the ethnic provenance of the food products that they buy: one club member, when giving me a tour of the supermarket by my new apartment, pointed out which ice cream brands were Tatar, and thus the ones that I should purchase; another club member brought me bottled water with a Tatar brand name to replace my bottle of *Raifskii Istochnik* ‘The Source of Raifa,’ water that is associated with a local monastery and considered holy by Christians. “Tatars don’t drink that brand of water,” he told me, “you shouldn’t buy

it.” By using a brand name in Russian with an Arabic-style script to signify that it is in fact a Tatar brand,¹² the producers of the liqueur¹³ are aiming for a potential customer base that includes both Russians and Tatars, particularly those Tatars who actively choose to purchase Tatar brands over competing Russian brands.

The final example of the symbolic use of Arabic orthography to be presented here can be seen in Figure 12, and is a piece of artwork drawn and lettered by Näjip Näkkash,¹⁴ a Tatar artist fluent in Arabic who has been instrumental in the return of Arabic calligraphy as an artform in post-Soviet Kazan – he also is the calligrapher for the edition of the newspaper *Iman* ‘Belief’ seen in Figure 10. The original of the artwork seen in Figure 12¹⁵ was displayed in a one-man show in the spring of 2001 in a gallery housed in a government building downtown; several government functionaries spoke at the show’s opening, including the Mufti of Tatarstan, and all but one of the ceremonial speeches were in Tatar. The works in the show included a variety of *tugra*’s, a art form where the graphic is composed of the Arabic letters of a person’s name, *shamail*’s, which are decorative religious plaques used as both art and (anti evil-eye) protection (even by not-particularly-religious Tatars), and illustrated poems. The illustrated poem shown in Figure 12 is Gabdulla Tukay’s *I Tugan Tel*, a poem that in song form is considered the unofficial Tatar anthem, and that is omnipresent in the public Tatar discourse of purity: not only is it the title of a monthly supplement in the newspaper *Mädäni Jomga* that is devoted to issues of language, but quotes from the poem (particularly the first stanza) are used as headlines for articles on linguistic purism, and can be found in the texts of the articles as well – quoting the poem automatically illustrates the

¹² These semiotics may not be perceived by Russian consumers.

¹³ *Bälzäm* is considered by many to be medicinal, and is not an alcohol product that is comparable with, say, beer or vodka.

¹⁴ A *nom de plume* – *näkkash* is an Arabic borrowing that means ‘artist’, ‘engraver.’

¹⁵ The work that was scanned for Figure 12 is a reproduction presented to me by the artist.

author's respectful and positive attitude toward the Tatar language. In present-day Tatarstan, it is the ultimate symbol of the emotional and cultural value of Tatar as a mother tongue, and as such bears quoting in its entirety:

I Tugan Tel

*I tugan tel, i matur tel, ätkäm-änkämng tele!
Dön"yada küp närsä beldem sin tugan tel arkili.*

*Ing elek bu tel belän änkäm bishektä köylägan,
Annari tönnär buyi äbläm khikäyat söylägan.*

*I tugan tel, härwakitta yardämeng belän sineng,
Kechkenädän anglashilgan shatügim, kaygim minem.*

*I tugan tel! Sindä bulgan ing elek kilgan dogam:
Yarlikagil, dip, üzem häm ätkäm-änkämne, khodam.*

O Native Language

O native language, o beautiful language, language of my dear father and mother!
I learned much in the world through you, my native language.

In the very beginning, with this language my dear mother sang lullabies,
Later, in the evenings, my grandmother would tell fairy tales.

O native language, your help is always with me,
Since my youth I have known that you are my joy and my sorrow.

O native language! It is in you that I made my first prayer:
Forgive, I said, me and my dear father and mother, my lord.



Figure 12. Gabdulla Tukay's poem I Tugan Tel, calligraphed and illustrated by Nəjib Nəkkash.

The text of the poem in Figure 12 is in the Tatar Arabic script, which is the orthography used by Tukay (1886-1913) at the turn of the last century, and is presented in its entirety as a central part of the graphical representation – the artist’s specialty is Arabic calligraphy. However, Näkkash seems to anticipate that the vast majority of his audience will not recognize the text of the poem, although they will possibly recognize the portrait of Tukay that sits above the text. The border of the poem, in lieu of a more traditional abstract or botanical graphical border, is devoted to excerpts of the poem in the Tatar Cyrillic script: the entire first stanza, the entire third stanza, and the first line of the last stanza. The text is so well known that it can be assumed that most of the audience can fill in the elided lines themselves.

In this work of art, Näkkash is using his Arabic calligraphic skills in an explicitly Tatar-specific way that is meant to be comprehensible to a general audience: his calligraphy is most often put to use in post-Soviet Tatarstan for classical Arabic words and religious texts, or for little-known art forms – the *shämail’s* are religious, the *tugra* is an art form little-known in Tatarstan, and its history and meaning is usually explained in articles about Näkkash and his work (e.g., Kayumov 1998), and most of his other work is related to religious calendars and books. Here, however, the Arabic calligraphy not only provides a fresh look at a poem that is ubiquitous to the point of triteness, but it also alludes to the pre-Revolutionary context of this poem, the era when this script was used; in particular, the 19th and early 20th centuries, generally considered to be the golden era of Tatar literary and cultural production. Words glorifying the Tatar language are raised to the level of art through the Arabic calligraphy, which also metaphorically links Arabic influence and the glory days of Tatar culture, before the Soviet-era Russification process began in earnest.

2.3.2 Lexical reform: The return of Arabic and Persian loanwords

Another language purification movement, found at both the individual and collective levels, is ongoing lexical reform by Tatar language planners and language professionals – less politically sensitive and less organized than orthographic reform, and not at all legislated. Soviet linguists, when engaged in the modernization and lexical development of the Soviet Union's more than 200 languages (under the aegis of Lenin's policy), worked towards a goal of transitioning monolingual speakers of minority languages to Russian bilingualism (Moskovich 1989). In the 1930s, Soviet linguists replaced most of the Arabic and Persian loanwords in Tatar with Russian loanwords, such that half of the entries in today's standard Tatar-Russian dictionaries (e.g., Ganiyev 2000) are Russian borrowings; in most cases the entries and the definitions are identical, e.g., the entry for the "Tatar" word *stena* [wall] gives the Russian definition *stena*. Present-day Tatar language planners have three main options for the development and modernization of Tatar: 1. return to the Arabic or Persian loanword that was used up until the 1930s; 2. attempt to find a relic form used before or contemporaneously with a Russian loanword; and 3. coin new terminology using native Turco-Tatar stock (for recommendations on this course of action see in particular Shamsutdinova 2000, Väliyev 2001, and Minhaj 1993).¹⁶ The course of action chosen most often has been the first one: a return to the Arabic and Persian loanwords of the past.

Islamic culture significantly influenced what is now Volga Tatar culture for over a thousand years, with economic and political ties established well before the early-10th century conversion to Islam of Volga Bulgaria. In the Islamic world, Arabic was the language of not only religion, but also literature, culture and politics, and the influence of Arabic on the

predecessors of the modern Tatar language was significant. The first Arabic loanwords date from as early as the 10th century, while Persian loanwords date from slightly later, as Persian became a widespread literary language only after approximately 1000 C.E. (Uli Schamiloglu, personal communication). As can be seen from the approximately 800 pages of Arabic loanwords listed in the 1993 reprinting of Mäkhmütov's dictionary of Arabic borrowings in Tatar, by the early 20th century a significant portion of the lexicon of written, literary Tatar was of Arabic origin – and the majority of these words were incomprehensible to those who were not educated in the literary languages of Arabic and Persian (a situation much like that of Ottoman Turkish vs. common Anatolian Turkish in the 19th century).¹⁷

In the 19th century, with the aim of closing this gap between the literary language and the language of the people, the scholar K. Nasıyri and other writers and poets started increasing the percentage of native Tatar words in their writing (Mäkhmütov 1993: 797). This process can be seen, for example, in the work of the poet Tukay, whose poem *I Tugan Tel* was discussed in the previous section. In Tukay's works written between 1905 and 1907, the Arabic-Persian element is 60-65 percent, between 1910 and 1913 it fell to just 25-30 percent of his writing. Even so, in poems such as his 1905 *I kaläm!* 'O, the pen!', the Arabic loanword content is as high as 65 percent (Mäkhmütov 1993, citing Ramazanov 1954). Therefore, in order to be comprehensible to the reading public, pre-Revolutionary Tatar texts are now presented with glossaries when they are reprinted in post-Soviet newspapers and magazines. For example, excerpts from the first issue of the journal *Sbura* printed in honor of the 90th anniversary of its founding in 1908 (in Mädriyeva 1998a) contain 177

¹⁶ These choices – retaining loanwords, coining neologisms, or finding archaic or dialectal variants – are quite common in language reform, particularly in post-colonial situations. See Richards (1989) and Heyd (1954) for parallels in Guatemalan Mayan and Anatolian Turkish, respectively.

¹⁷ In the Arabic-Persian stratum, the Persian element was approximately 10-12% (Mäkhmütov 1993).

words, 32 of which are translated in the glossary at the end – without this glossary 18% of these excerpts would presumably be incomprehensible to the modern Tatar reader.

This lack of comprehensibility of texts written not even one hundred years ago is due, of course, to the previously mentioned language engineering that took place from the 1930s onward, where the Arabic-Persian stratum of the lexicon was mostly replaced with Russian borrowings (many of which are themselves borrowings from Latin and modern European languages). This process differed from the de-Arabicization movement of the late 19th century in two major ways: first, the 19th century movement was grass-roots, while the Soviet-era language reform was institutional and top-down, and second, the 19th century movement had as a goal the increased usage of words of native Tatar and Turkic stock, while the Soviet-era reform had an end result of a Tatar lexicon that is half Russian. It is believed by many that religious concerns were a deciding factor in this de-Arabicization reform:

“Great attention should be paid now to the Tatar language, from whose lexicon were discarded thousands of words and expressions of Arabic origin. Ignorant bureaucrats associated all that was Arabic with Islam, and not with the rich Muslim culture that exerted an enormous (at that time, progressive) influence in ancient times on the language of the Tatar people, the development of their culture, their writing system, book making, and education” (Mäkhmütov 1993: 8).

This relexification of Tatar, in combination with two alphabet changes in quick succession (again, from Arabic graphics to the Latin-based *Yangalif* in 1928, and then in 1938 from *Yangalif* to the modified Cyrillic script still in use today), caused a significant discontinuity with pre-Revolutionary Tatar culture, perceived by some to be purposeful and discriminatory:

“Tatars used an Arabic writing system for almost a thousand years... But in the 1920s-30s, the national culture suffered a great loss: the alphabet was changed twice without the consent of the people, leaving the older generation illiterate in their native tongue, and the younger generation cut off from the rich literature of the past. For many years religious literature was prohibited, literature that the

population reached out for thirstily, feeling a spiritual devastation after almost a century of persecution of the Muslim culture” (Mäkhmütov 1993: 8).

While the use of Arabic and Persian loanwords decreased from the 1930s onward, they started appearing once more in the language of the press around the time of *perestroika* (the mid-1980s to 1990), a gesture readily perceived by many Tatars as more than purely linguistic in nature. For example, Safiullina and Fyodorova (2000) explicitly link the return of Arabic-Persian loanwords to “the start of democratic changes in the country, due to the influence of the awakening of the people’s national self-consciousness.” They interpret the increased use of archaic Arabic and Persian words as related to “the democratization of society, the revitalization of Islamic observance, the opening of Muslim schools, increased relations with Arabic countries and Turkey, and favorable conditions in our country’s socio-political state” (ibid.).

Arabic and Persian loanwords in Tatar can be separated into two groups: everyday words that are encountered in regular speech, and more cultured words that are used most frequently in literary registers. Words of the first sort, everyday words unmarked for register, were for the most part not removed during the Soviet-era relexification process and are generally felt to be native. “...[I]t is often possible that in a language this or that concept can only be expressed with a loanword. For example, in the Tatar language there are no words to express concepts such as *kitap* [book], *däftär* [notebook], *mäktäp* [school], *tärtip* [upbringing], and *tarikb* [history]. They can only be expressed with these loanwords (which have come from Arabic)...And we quite often do not feel that these words are borrowings...” (Safiullina and Fyodorova 2000). Mäkhmütov, writing in 1993 in the introduction to his Arabic loanwords dictionary, expresses the same sentiment, using many of the same examples: “This stratum of borrowings...is found not only in literary language but also in

the conversational language of the Tatar population, who perceive this lexicon as their own, as Tatar. Who would say that the words *kitap* [book], *däftär* [notebook], *mäktäp* [school], *iman* [belief], *sabiy* [infant], *sabir* [patience]... and thousands of similar words are not Tatar?” (1993: 8).

The majority of the archaic Arabic words that are in the process of being revived in post-Soviet lexical reform, however, are more literary or formal words, words used to describe politics, literature, culture, and religion. Their use is not yet standardized – for example, one can find both *revolutsiya* ‘revolution’ and its Arabic equivalent *inkiyälab* in one and the same article (for example, in Safiullina and Fyodorova 2000, which is itself in part about the inconsistent use of Arabic loanwords) – and more importantly, their use is not uncontroversial. Some Tatars resist the idea of replacing Russian words that are now standard in Tatar with Arabic equivalents: in a response to the suggestion that word *shyot* ‘gathering, rally’ be removed from the lexicon, one author argues in favor of keeping it – if he “forgot” the word, he claims, he would lose associations with much of his childhood and teenage years. He writes, “Let’s not throw out harmonious and meaningful words from our language that have been used for centuries, and fence in and narrow our interactions, impoverish our language. Let’s stay ourselves, and speak with meaningful words understood by us in our own country” (Asarov 1999). However, sentiments so overtly in favor of retaining Russian loanwords are rarely heard in post-Soviet Tatar discourse. At the other extreme, we find stances such as the one expressed in an article entitled *Bäylänçhek süzlärdän ariniyek!* ‘Let’s get rid of intrusive words!’, where the author has a hope that “in time every simple person will live understanding a Tatar language where Arabic and Persian words are heard” (Kotlikäläm 1997). This author (whose name, which translates to ‘happy pen’, is clearly a pseudonym) deems the Russian words in the Tatar lexicon “parasites,” and suggests,

among others, the following re-replacements, all archaic Arabic loanwords: *järidä* ‘newspaper’, *mäjällä* ‘journal’, *kbojjät* ‘document’, *sänäd* ‘argument’, *shäkel* ‘form’, *mokhajir* ‘emigrant’, and *teg’mir* ‘repairs’ (ibid.).

Kotlikäläm’s article generated several responses, all published in the same paper within the next several months, where the authors advocated a more moderate path in lexical reform. Some philologists would like to differentiate between words that are etymologically Russian and words that have an “international” origin. In response to Kotlikäläm suggestions, Faziljanov (1997) points out that many the words he had labelled as Russian were actually from Latin (e.g., *orator*, *dokument*, *etika*, *emigrant*), Greek (e.g., *tema*, *genotsid*), and French (e.g., *remont*). Additionally, Faziljanov finds the suggested replacements to be archaic and inappropriate, and questions whether their use is “better” or “cleaner.” Faziljanov’s opinion is that archaic Arabic loanwords are “incomprehensible,” and although he concedes that “after being repeated sufficiently often in the press, words like *jömhüriyat* [republic], *ik’tisad* [economics], *mädäniyat* [culture], *firka* [group], *säyäsi* [political], *täräkkäyat* [development], *mökhabirir* [editor]...and other words have become almost customary,” he believes that “it will be a good long time before they really enter the spoken language” (ibid.).

Other philologists believe, probably correctly, that the average speaker is unaware of the etymologies of most of the words in question, and generally unable to differentiate between words of international origin and those of Russian origin. One such philologist writes:

“...many terms are words taken from Greek and Latin. But they are not perceived as such. So they seem to us to be entirely Russian. For example: *protsess* [process], *apparat* [apparatus], *absoliut* [absolute], *avtomat* [automatic machine, vending machine], *informatsiia* [information] etc. – who wouldn’t say that these are Russian words?

And having become accustomed to hearing such words...who wouldn't want to study in Russian?"(Nogmani 1991).

His suggestion is "to translate scientific and technological terms into the mother tongue" (ibid.). This preference for using native stock, "Tatar's own resources," is one commonly voiced in discourse on lexical reform, where suggested replacements include "forgotten Turkic words" such as *yazgın* 'secretary' lieu of *sekretar*, and *kiarkhanä* 'factory' in lieu of *fabrika* and *zavod*. (Khäyrullin 2001).

Another option in lexical reform, advocated by some Tatar linguists, is to use all available resources. For example, one philologist, writing on the creation of new Tatar medical terminology, suggests: "It is necessary to maximally use not only sources of a literary nature, but also dialects and regional varieties of borrowings, new forms, and neologisms, thus replenishing the lexical contents of the language and reactivating processes of word-formation" (Shamsutdinova 2001: 26).

However, the most common course of action has been the return to selected Arabic loanwords, portrayed by supporters as words that were, "in essence, forfeited in the process of Sovietization of the national culture" (Mäkhmutov 1993: 9). A great many of the Russian nouns (and these loanwords are almost entirely nouns) used in political and cultural discourse have been replaced in the post-Soviet daily press by their pre-Revolutionary Arabic equivalents: for example, in the realm of politics we find the words *khakimiyat* 'ruling power', *säyäsat* 'politics', *ijtimagiy* 'social', *möstäkil'lek* 'independence', and many more.

The words listed above are quite commonly used in the post-Soviet press, far more often than their formerly-standard Russian equivalents, and their usage is reasonably uncontroversial. However, a complaint that surfaces rather often is that the press is going too far in their use of Arabic words, that there are no established norms, and that words are

used inconsistently within the same paper: “the language of the press is moving away from the language of the people. Simple people have begun to say that there is no sense, that they don’t understand anything that is written in newspapers and magazines” (Jälälova 1996). Complaints about the increasing gap between the language of the press and the language of the people can be found in both articles by philologists and other language specialists and in letters to the editor from ordinary citizens. For example, it is a philologist who writes that:

“it has become more difficult to read Tatar newspapers and journals from one end to another in one sitting. After picking up a newspaper and starting to read, you must stop and busy yourself with foreign words that are meaningless in Tatar, and then you remain in thought...” (Väliyev 1999).

The use of incomprehensible Arabic loanwords is explicitly related to the construction of Tatar identity, as can be seen in the title of this letter to the editor, (most likely ironically) given the title *Min Tatar tügel, akbrisi* ‘I am not a Tatar, it would seem’:

“In the Tatar newspapers and journals that are published these days, I continually encounter words that I just do not know. For example: *näfasät* [moment], *manzara* [view, spectacle], *wäzgiyat* [situation], *ik’isad* [economics], and others. They aren’t in any dictionary... Also, if there are Tatar terms available, skip the Russian words, for example, *problema* [problem], *predpriyatie* [enterprise, business], *faktor* [factor], *sotsiologia* [sociology] and others. All right, there aren’t translations for words like *kibernetika* [cybernetics] and *kompüter* [computer], but for the Russian words written above, there are Tatar terms after all” (Iskändärova 2001).

Note that the letter writer does not distinguish between Russian words that are of native stock (*predpriyatie*) and Russian words that are themselves borrowings (*problema*, *faktor*, and *sotsiologia*): she labels them all simply “Russian.” Interestingly, the Tatar terms that would be used in place of these Russian terms would either be Arabic loanwords (e.g., *mäs’älä* ‘problem’) or do not exist (e.g., there is currently no native equivalent for *faktor*) and thus would need to be neologisms. But Iskändärova is indeed correct when she states that the Arabic words listed above cannot be found in most Tatar dictionaries, with the exception of

special dictionaries of Arabic borrowings, dictionaries that are not commonly available. The first few times that an archaic Arabic loanword is used in a present-day newspaper or journal, it will usually have a translation adjacent to it, either the Russian loanword it is replacing or an explanatory Tatar phrase, for example: “Bälkem, **motlak** (absoliut) tigezlek te tügelder, läkin tigezlek printsibi tiesh” ‘Maybe it is not **absolute** (absolute) equality, but there should be a principle of equality’ (Excerpt taken from the first issue of the post-Soviet Tatar journal *Idel*, cited in Safiullina and Fyodorova 2000). However, readers who miss the first few times an archaic Arabic loanword is translated or explained will be left without a means for understanding it, unless they have access to an Arabic dictionary.¹⁸

Additionally, due in part to semantic fields that do not entirely overlap, or to mismatched polysemy, some of these returning Arabic loanwords are not entirely synonymous with their Russian “equivalents,” or are being used differently than they were in the past. For example, the word *mökbarrir* was used in pre-Revolutionary times to refer to the professions of writer, journalist, author, and editor, but in post-Soviet Tatar is used only as a substitute for *redaktor* ‘editor.’ Väliyev (1999) notes that there is a semantic mismatch that causes *mädäniyat* ‘culture’ to be used inappropriately as a substitute for Russian *kul’tura*: apparently the Arabic word has a spiritual component that makes expressions like *matdi mädäniyat* ‘material culture’ inappropriate (See Faller 2000 for more on the semantic mismatches and particularized uses for these words). And Minhaj (1997) does not find *mäs’älä* ‘problem’ to always be an appropriate substitute for *problema* – semantic nuances differentiate the two, such that they are not sufficiently synonymous to be used

¹⁸ These dictionaries of Arabic borrowings in Tatar are published in small quantities and are quite difficult to find.

interchangeably. He suggests, “Let both the words *problema* and *mäs’älä* be used. The use of words with partially overlapping meanings shows a language’s richness” (for more on the perception that increasing the number of synonyms is equivalent to increasing Tatar’s linguistic resources see Madriyeva 1998b). Due to a lack of either standardization or widely accepted norms, the simultaneous use of both Arabic and Russian loanwords is currently the *de facto* norm. Lexical collocations such as *ijtimagïy problemalar* ‘social problems’, where the adjective is a formerly archaic Arabic loanword, the noun is a Russian borrowing of “international” origin, and the plural suffix is Tatar, are emblematic of post-Soviet Tatar linguistic reality.

2.3.3 A comparison with 20th century Turkish language reform

The connection between linguistic purism and nationalist projects is well documented, especially for the case of 20th century Turkey. A comparison between Tatar and Turkish nationalism, linguistic purism, and language reform is particularly illuminating both because of the similarities in the linguistic history of these two Turkic languages, and because it demonstrates quite clearly the effect that macro-social structure and forces can have on linguistic structure: due to their different socio-political and cultural situations, the language reforms in Tatarstan and Turkey have had different results, particularly with regards to Arabic and Persian loanwords.

The ancestors of the modern Turks converted to Islam in the beginning of the 11th century (approximately 75 years after the conversion to Islam of the Bolgars of the Middle Volga region), and lexical borrowing from Arabic and Persian began in the Seljuk empire (1040-1157) (Lewis 1999: 5). The use of what is known as “Ottoman Turkish” as a literary language dates back to the 13th century, at which point there were already many Arabic and

Persian loanwords in the literary high language (Heyd 1954: 9). Ottoman Turkish was both a literary and an administrative language, distinct from the vernacular of the people and incomprehensible to them, as it required competence in both Arabic and Persian. In fact, “every Persian and every Arabic word was a possible Ottoman word. In thus borrowing material from the two classical languages a writer was quite unrestricted save by his own taste and the limit of his knowledge...” (Gibb 1900-9 cited in Lewis 1999:7) While the Persian stratum of the Ottoman Turkish lexicon was large, the Arabic stratum was even larger, because entire families of words based on a trilateral root were borrowed en masse, along with the morphological rules for their derivation (Lewis 1999: 6). By the 16th century, Ottoman Turkish had been so heavily relexified with Arabic and Persian loanwords, and used so many foreign grammatical morphemes and constructions, that the native Turkic elements in a sentence could be limited to suffixes or the copulas *di* or *dır*, or there could be no native Turkish elements *at all* – for example, Lewis (1999: 8) presents three couplets of a 16th century Ottoman Turkish ode that contain not one syllable of Turkish. This heavy reliance upon foreign words and grammar continued into the 20th century, as can be seen in following story, related by a Turk who is writing about his early education in Damascus in 1908, the year that the 1876 constitution was restored in Turkey:

“The Arabs suddenly started on nationalism and took to making fun of Turkish. One day in the classroom we saw a half a dozen or so lines written on the blackboard, headed ‘What is the Turkish language?’ We read the writing to ourselves; it contained not a single word of Turkish. Written in conformity with the style and rules of Ottoman, it ended with *-dir*. The Arabs had repeated this suffix several times, underlining this string of *-dir*s and writing in front of it ‘Turkish is this. That is to say, it’s *dirdir* [tedious babble].’ That day we four or five Turkish pupils very nearly came to blows with a whole class, and became devotees of Turkish from that day on” (Tankut 1963: 113, cited in Lewis 1999: 40).

In fact, 40% of a 1907 Ottoman-Turkish conversation grammar is devoted to Persian and Arabic grammatical rules (Lewis 1999: 16), which included derivational morphology for the creation of adjectives and nouns. Other grammatical rules that were imported along with the Persian and Arabic loanwords included agreement for gender (Persian and Arabic have gender; Turkish does not) and the use of Persian *izafet* to connect nouns and qualifiers (Lewis 1999: 6-7). Persian *izafet* should not be confused with native Turkic *izafet*, which, although it is also used to connect words (*izafet* < Arabic *idāfa* ‘attachment’), is a different morphosyntactic construction.

Calls for language reform started in the mid-19th century, and are analyzed by some as due to increased European influence and the rise of a Westernized middle class in Turkey (Heyd 1954: 10); these 19th century reforms were known as the Tanzimat reforms, and were isolated, mostly individual attempts with little-to-no effect (Lewis 1999: 2). The next wave of language reform came with the Young Turks (mostly in the years 1908-9), and were associated with their nationalist and pan-Turkic goals – it was realized that in order to “arouse political consciousness in the masses” the language of the press would be of utmost importance and would thus need to be intelligible to all, even the uneducated (Heyd 1954: 16). However, it was not until the establishment in the 1920s of the rule of Mustafa Kemal, better known as Atatürk ‘Father Turk’, that both massive and officially sanctioned reform of Ottoman Turkish began. Atatürk’s stated political ideology called for the creation of a nationalist, secular, populist, and revolutionary republic (Heyd 1954: 19), where political independence would be mirrored by linguistic independence: in the foreword to a 1930 book, he wrote, “The Turkish nation, which is well able to protect its territory and its sublime independence must also liberate its language from the yoke of foreign languages” (Lewis 1999: 42). In the new climate of nationalism, along with the new westward

orientation of Turkey, Arabic and Persian loanwords were perceived as a “national disgrace” (Heyd 1954:19) that had no place in the language of the new secular nation. Linguistic and religious reform went hand-in-hand: 1928, known as the year of *dil inkıyabı* ‘the language revolution’¹⁹ is also the year that it was decreed that Islam was no longer the official state religion, and the study of Arabic and Persian as foreign languages was removed from the standard curriculum for secondary-level education in 1929. Other language-related religious mandates, such as the 1928 declaration that the Friday sermon in mosques was to be in Turkish only, and the 1932 declaration that the call to prayer was also to be in Turkish only, were short-lived (Heyd 1954; Lewis 1999).

A change in orthography played a major role in Atatürk’s Turkish language reform. The move in 1928 from an Arabic-based to a Latin-based alphabet has been interpreted as a gesture designed both to break ties with the Islamic East and to facilitate communication with the West (Lewis 1999: 27; Anderson 1987: 45-6). The Ottoman Turkish alphabet, composed of Arabic letters and three Persian-specific letters, was well-suited for writing Arabic and Persian words, but ill-suited for writing Turkish words.²⁰ Some symbols represented sounds not found in native Turkish words, while other symbols were used for several phones, causing potential ambiguity. For example, there were three vowels in the Ottoman Turkish alphabet but eight phonemic vowels in Turkish, such that the combination of letters *ʔw/w*, the transliterated equivalent of the Arabic letters, could be read as the following words: *ulu* ‘great’, *ulu* ‘possessors (Arabic)’, *ölü* ‘dead’, *evli* ‘married’, *avlu* ‘courtyard’, and finally, *acli* ‘stocked with game’ (Lewis 1999: 27); the reader would need to know not only the context of an ambiguously spelled words, but also all relevant Turkish, Persian and

¹⁹ After this revolution, it was known as *dil devrimi*, still ‘the language revolution’, but without the Arabic loanword *inkıyap*.

Arabic options in order to disambiguate. As a result, Turkish words spelled in the Arabic-based alphabet were often not as easily comprehensible as their Arabic or Persian synonyms, which would be spelled unambiguously (*ibid.* 28).

There are some parallels between the Ottoman Turkish Arabic alphabet and the modified Cyrillic alphabet used for Tatar, which is better suited for the Russian language for which it was designed than it is for Tatar. The Tatar Cyrillic alphabet too has symbols for sounds that are not found in the native lexicon, for example, ‘ж’ /ʒ/ and ‘ц’ /ts/. And, like the Turkish Arabic alphabet, the Tatar Cyrillic alphabet has several symbols that represent more than one phone and can be read ambiguously. For example, Cyrillic ‘е’ when found word-initially represents Tatar [ji] and [je], which are contrastive and can be found in minimal pairs, while between consonants it represents Tatar [e]. And Cyrillic ‘я’ is used for contrastive Tatar [ja] and [jæ] both: in most, but not all, ambiguous words, the front version is marked by a “soft sign,” ‘ь’, while the back version is marked by a “hard sign,” ‘ѣ’. Additionally, Cyrillic ‘в’ is used to represent [w] in Tatar words and [v] in Russian loanwords ([v] is not in the native phonetic inventory); readers must therefore know a word’s provenance in order to read it correctly.

The new Latin alphabet designed for Tatar removes all of these ambiguities, and represents Tatar words and Russian words equally well. The Latin alphabet developed for Turkish in 1928 (upon which the new Tatar alphabet is based) also removed the orthographic problems that gave preference to foreign loanwords, and turned the tables, as it were. While Arabic characters were well-suited for Arabic and Persian words, and not for Turkish words, making the Arabic and Persian words seem “natural,” the new Latin-based

²⁰ The same is true, to a lesser extent, for the modified Arabic script used for Tatar until the 20th century.

alphabet caused the Arabic and Persian loanwords to look “distinctly foreign”: they clearly violated the rules of vowel harmony, and the etymology of and rules governing the inflection and derivation of words based on a single trilateral root became unintelligible rather than transparent (Heyd 1954: 23).

The lexical reform of the Turkish “language revolution” focused on removing these Arabic and Persian loanwords and replacing them with “appropriate” equivalents. The process was originally populist in nature, with lexical collection committees set up throughout the republic, committees that managed to gather over 125,000 forms (where 1 form = 1 suggested replacement word) in the space of just one year (Heyd 1954: 26-7). Additionally, in the early days of language reform, daily newspapers would publish Arabic and Persian loanwords on their front pages and request native equivalents from their readers. It was discovered that the vernacular and dialects of Turkish could provide native equivalents for concrete terms, but not for abstract ones (ibid. 29). The language committee turned to alternative sources for acceptable words: remote Anatolian dialects, Turkic languages spoken outside of Anatolia, pre-Islamic manuscripts, and neologisms (ibid. 31). Lewis (1999: 70) contends that Atatürk’s intent was that Arabic and Persian *technical* terms be replaced with native equivalents, but that the Language Society established under his rule and with his encouragement overstepped its bounds and also replaced standard lexical items found in vernacular Turkish as well as in the high language.

Here in the Turkish case we once again find linguistic purism with “selective targeting”: Turkish was to be purified of Persian and Arabic loanwords, but in addition to the native and semi-native lexical sources listed above, European loanwords were also considered to be acceptable substitutes. In 1935, the Secretary General of the Language Society explained this policy, saying that “words like *katib*, *müdir*, etc. were relics of a bygone

era. At a time when the Turks were adopting Occidental civilization in its entirety, the Western equivalents of such terms should be preferred” (Heyd 1954: 77). Other language reformers agreed: “Mark this well: the thrust of the reform movement is specifically against Arabic. Arabic words have to be discarded come what may, for this is a generation that is fed up with the domination of Arabic. If the French equivalent were to replace it, that’s fine” (Banguoğlu 1987: 303 in Lewis 1999: 118).

Many neologisms coined by language reformers (neologisms that often went against the rules of Turkish grammar) have phonetic forms or spellings that “fortuitously” resemble their English or French equivalents, even though these neologisms were always presented with some sort of (specious) Turkish etymology. For example, *okul* ‘school’ (compare French *école*) is ostensibly derived from the verb *okumak* ‘to read, to study’, which would be reasonable but for the fact that there is no deverbal nominalizing suffix in Turkish that resembles *-l* in any way. The etymology given for *genel* ‘general’ was Turkish *gen* ‘wide’ + the adjectival suffix *-el*, imported from French. The neologism *imge* ‘image’ was explained as Old Turkish *im* ‘password’ + the unproductive nominal suffix *-ge*, although it is clearly based on the spelling of English and French *image*, and *terim* ‘term’ was claimed to be derived from Old Turkish *derim* ‘assembly, gathering’ rather than based on French *terme* or English *term*, although apparently no attempt was made to explain either the consonantal devoicing or the semantic shift (Lewis 1999: 94ff).

In addition to resurrecting unproductive Turkish morphology (e.g., *-KI* and *-It* as nominalizers), and importing morphology from other Turkic languages (e.g., *-(E)v*, the suffix for verbal nouns in Bashkir and Kazakh), Turkish language reformers imported English and French derivational morphology. The French adjectival suffix *-el* found in the loanword *kültürel* ‘cultural’ was transformed into the “pure” Turkish adjectival suffix *-el* by using

several s-final nouns as analogical base forms. English *-man* became the agentive Turkish suffix *-mAn*, now found in such standard words as *öğretmen* ‘teacher’ (*öğret-* ‘teach’ + *-mAn*) (ibid. 94ff). One of the more spectacular neologisms is *ulusal*, the replacement for the Arabic loanword *milli* ‘national’: *ulus* ‘nation’ was taken from Mongolian, and the adjectival suffix was borrowed from French (ibid. 102): therefore, the post-lexical reform “pure” Turkish replacement for the Arabic adjective meaning ‘national’ is actually half-French and half-Mongolian.

The end result of 20th century lexical reform in Turkish, which continued up until the 1980s, is that not only is Ottoman Turkish incomprehensible to speakers of modern Turkish who have not studied it specially, but also texts written during the 1930s, the early days of the purification movement, are also incomprehensible, and apparently this unintelligibility is considered to be “normal” (Lewis 1999: 142-3); for example, the Turkish constitution has been translated and retranslated into modern Turkish several times. Arabic and Persian words were expelled inconsistently – for example, *hakim* ‘judge’ was replaced by Turkic *yargıç*, but *mahkeme* ‘court’ remained (Heyd 1954: 43) – and a significant percentage of the lexicon of written Turkish is still comprised of Arabic and Persian loanwords. While the percentage of the population literate in Turkish has risen significantly, from 9% in 1924 to 82% in 1995 (Lewis 1999: 37), there still remains a gap between the language of the state and the language of the people. Lewis contends that the language reform has had little effect on the speech of the “common people,” who still speak the “old language” (which had Arabic and Persian loanwords, but to a far lesser extent, and did not have foreign grammatical rules) (1999: 141); and because the Turkish Language Society replaced so many Arabic and Persian words with neologisms and borrowings, even non-technical and everyday terms, the *öztürkece*

‘pure Turkish’ that is used as the administrative language and language of the intelligentsia is still incomprehensible all but the educated elite:

“...many critics were of the opinion that the Society had defeated its own purpose. Instead of developing the existing language so that it would be understood even by the common people, the Society was said to have created a new artificial language, an official and school language very different from the language of ordinary conversation and as unfamiliar and unintelligible to the masses as the old Ottoman had been. A situation had arisen in which parents no longer understood the language of their children, nor the public the language of the authorities. Thus a new and dangerous cleavage had been created between the intelligentsia and the masses, and the gap between the written and the spoken language had widened again” (Heyd 1954: 47).

Neologisms were formerly introduced in two ways, in schools and via the press, but now they are no longer taught in schools, and when newspaper readers do not know the meaning of a word they encounter in the paper, they make *ad hoc* interpretations (Lewis 1999: 144-5). As with some of the archaic Arabic loanwords being brought back into usage in post-Soviet Tatar, post-Revolutionary Turkish neologisms often do not precisely match either the semantics or the polysemy patterns of the words they are replacing; for example, *tejrübe etmek* meant ‘to experiment’ and ‘to test’, and ‘to experience’, while its replacements *denemek* means only ‘to experiment’ (Heyd 1954: 75). Turkish writers and older speakers complain of the lack of synonymic richness and lexical resources in the new languages, and Lewis, who is decidedly critical of the language reform in general²¹ contends that it “left the Turks with virtually no choice of levels of discourse” (1999: 144).

The similarities and differences between the Turkish and Tatar language reforms can be seen as grounded in the larger socio-cultural and political context. Both reforms are sited within a nationalist project, and the selected targets of the reforms are languages perceived as threats: the Turkish focus was on the purification of Arabic and Persian influence, and

European linguistic influences were deemed to be part of “pure” Turkish, while the Tatar focus is on the purification of Russian influence, and Arabic (and to a lesser extent Persian) is deemed to be a part of “pure” Tatar. While the Arabicization of Turkish was done by the Turks themselves, the Russification of Tatar was brought about not by Tatars but by an imperial (in the sense of the Soviet Empire) government. Additionally, the “break with the past” in post-revolutionary Turkey included a move towards secularism, while the Tatar break with the Soviet past takes the form of a move towards an Islamic identification, albeit not always congruent with religious observance. As a result, the purification movements of these two Turkic languages have had different results with regard to their Arabic loanwords: Arabic influence was rejected and expelled by the Turks as having no place in the post-Ottoman Turkish identity, but is currently being embraced by many Tatars as part of the post-Soviet Tatar identity.

2.4 Conclusions

Linguistic purism and language reform in present-day Tatarstan are best explained by means of a linguistic-anthropological correlation of language as a practice and language as a system, and also by locating linguistic practice within a larger socio-cultural and political context. Language in Tatarstan is used as a symbolic resource, and an examination of Tatar language ideologies shows the link between post-Soviet Tatar social structure and the post-Soviet Tatar linguistic system. Ideologies are reproduced by practices, and in Tatarstan the language ideology that I am calling the “discourse of purity” is found in both individual linguistic practice and in communal linguistic practice, most notably in the pages of the daily press. The Tatar discourse of purity is a counter-hegemonic discourse that is set in the post-Soviet

²¹ This is clear from the title of his book: *The Turkish language reform: A catastrophic success.*

context of increased Tatar political power and autonomy that is nonetheless situated in a centralized and Russian-dominated Federation. Linguistic purism can be interpreted as an attempt to both add value to and control the Tatar language, which in the logic of language ideologies is seen as a part of a feature cluster that defines the boundaries of the Tatar people, and as a metonymic representative of the Tatar nation. Purification is not only a response to the “impure” Tatar (grammatically incorrect, misspelled, mistranslated Tatar that may also have phonetic interference) that is omnipresent in urban Tatarstan, it is also working towards the establishment of a realm or style that is purely Tatar and thus a counterpart to the pure Russian realms that are guaranteed by the existence in Tatarstan of Russian monolinguals.

Tatar linguistic purism takes several forms: on the individual level there is style shifting that is characterized in part by verbal hygiene, and on the communal level in organized and sometimes officially legislated purification movements, in particular orthographic and lexical reform. These purification movements are characterized both by a logic of oppositional identity and by selective targeting of only linguistic influences from the language and culture that is perceived as threat; here, the Russian language and culture. Barbara Einhorn writes of post-Soviet East Central Europe: “The search for untarnished values and identities has leapfrogged the often unpleasant realities of both state socialist and Second World War history, turning instead to the spirit of nineteenth-century...nationalism” (1992: 45). While 19th century Tatar nationalism was actually more along the lines of pan-Turkic solidarity and separatism (Frank 1998), the use of Arabic loanwords (and the periodic symbolic use of Arabic orthography) does appear to be a partial attempt to skip back in time over the Soviet era and join up once more with the pre-Revolutionary Tatar culture that was left behind, a culture that was part of a larger, international Islamic cultural system.

In present-day Tatarstan, there is a conflation by some of Tatar and Muslim identity, as exemplified by this exchange, witnessed at the central market in Kazan, the *Kolkhoznyi Rynok*. A shopper, having read the nametag of a salesclerk, is surprised that she is not Russian and enquires, “Are you a Tatar?,” to which she replies, “Yes, I’m a Muslim.” In fact, even the words used to describe the two ideals of Tatar linguistic purity, *saf* ‘pure’ and *ädabi* ‘literary’, are themselves Arabic loanwords, although this fact may not be known to the average Tatar speaker. The semiotic functions of metaphor and metonymy link both orthographies and lexical borrowings to the times and cultures that are their sources of origin: for example, runic writing is linked with both pan-Turkic sentiment and theories of ethnogenesis that highlight the antiquity of the Tatar nation, while the new Latin-based orthography can be read as a symbol of cultural, political, and economic alignment with both Turkey and the West. Language purification movements are a common component of the cultural redefinition that often comes after significant political change, and the print language of the press often plays an important role in both this linguistic and cultural redefinition. The post-Soviet Tatar-language press is a crucial part of the process of linguistic and cultural purification, and is a two-pronged means of reform: it presents discourse on the topic of purity and purification, and also leads by example, both by presenting actual linguistic models of new-old Arabic loanwords in use and by being a model of lexical reform in action. While the embrace of the Tatar-language press of Arabic loanwords is not completely uncontroversial, it is, to a certain extent, a *fait accompli*, and also consistent with the communal creation of a post-Soviet Tatar identity that is buttressed by the return of Islamic observance and by Tatarstan’s increased religious, educational, and socio-economic ties with Arabic countries. Linguistic purification and de-Russification are bound up in the Tatars’ ongoing struggle to resist religious, cultural, and linguistic assimilation into the Russian

majority. Tatar discourse on language is often either symbolic of, or explicitly related to, discourse on nationhood, with the hope that through the preservation of the integrity and distinctiveness of the language, the integrity and distinctiveness of the nation can also be retained. As the saying goes: *Tugan teldä – millät yazmışi* ‘In a native language is the fate of a nation.’

Chapter Three

The code-mixing of Russian words in “Tatar-preferred” style

3.1 Introduction

In discussions of endangered languages, three mechanisms of language attrition are often described: (1) the contraction of functional domains, also known as *encroaching diglossia* or *leaky diglossia* (cf. Dorian 1981; Gal 1979; and Grenoble and Whaley 1998, *inter alia*), (2) the shrinkage of stylistic options, including loss of register distinctions (cf. Dorian 1994; King 1989; and Taylor 1989, *inter alia*), and (3) a change in code-switching patterns such that the dominant language becomes the ‘matrix’ language into which content words from the contracting language are ‘embedded’ (Myers-Scotton 1993, 1998; Fuller 1996). While the first two of these mechanisms of language attrition can be found to a certain extent in the Tatar spoken in present-day Tatarstan, in this chapter I will identify another pathway of language attrition, one that has not yet been described and is not accounted for by any current model of language attrition or language shift: the unconscious code-mixing of dominant-language discourse-pragmatic words in minority-language speech that is otherwise free of code-switching. In this chapter I will examine specifically the code-mixing of Russian words in Tatar speech, and in Chapter Four I will demonstrate that both code-mixing and discourse-pragmatics are central to understanding both the lexical and structural borrowing found in language contact situations sociolinguistically similar to that of Tatar and Russian.

Studies of language obsolescence commonly cite encroaching diglossia as the most obvious sign of language attrition. In Chapter One I described the asymmetry of functional domains for Tatar and Russian in present-day Tatarstan, and noted how Russian appears to

be continuing in its encroachment upon Tatar domains despite promotive language policies on the part of the Tatarstani government. “Leakage” between language variety boundaries in diglossic situations was first noted by Ferguson (1959), and later scholarship adopted the term *leaky diglossia*, while the term *encroaching diglossia* was first used by Dimmendaal (1989) to describe language shift in East Africa. Along with a contracting population of speakers, languages in the process of a multi-generational shift have contracting functional domains, and often become limited to domestic contexts only. Then when there is a gap in transmission and parents choose to transmit only the dominant language or children choose to respond only in the dominant language, the final functional domain of the minority language is lost, and it becomes a language that is no longer spoken at all. Examples of contracting functional domains are ubiquitous in case studies of endangered languages; some of the most clear-cut examples include Nubian in Egypt, Scottish and Irish Gaelic, Norwegian in America, and Breton in France. For example, “in order to participate in the full range of daily-life activities in both urban and non-urban areas...Egyptian Nubians have to use Arabic. Thus, Arabic is taking over function by function, while Nubian is becoming more specialized to topics related to family occasions” (Rouchdy 1989: 100). The main causes of the increasing encroachment of Arabic include a resettlement (due to the construction of dams) that disrupted traditional social and economic structures and caused increased contact with monolingual Arabic speakers in villages and cities, Arabic newspapers from Cairo that are imported into villages, an increase in intermarriage, an increase in the number of educated women (who tend to address their children in Arabic rather than Nubian), and the fact that as a religious language Arabic is necessary for integration into Egyptian Muslim society (ibid. 92-96). The traditional domains of Scottish and Irish Gaelic

have been eroding as well, although for a much greater period of time than found in the Nubian case, where the “tip” towards Arabic was notable for its rapidity. In both Ireland and Scotland, religious services are shifting away from Gaelic, traditional medicine is being replaced by national health care, leading to the loss of botanical medical knowledge and vocabulary, and traditional rural professions (fishing, weaving, farming) have been disrupted by socio-economic changes, such that the registers and vocabularies associated with these domains are being lost. In addition to the domestic sphere (where transmission at this point is often self-consciously part of the revitalization process), Gaelic has retained its traditional cultural domains of storytelling and singing (Watson 1989: 51-54).

The “tip” towards English for Norwegian in America seems to have been predicated by World War I and its dramatic rise in anti-foreigner (and anti-foreign language) sentiment. While up until the war Norwegian had a variety of traditional domains, including the home, agriculture, church, summer schools, immigrant newspapers and journals, choral groups, and Norwegian societies, with the advent of the war English began encroaching quite rapidly into these Norwegian domains. Norwegian-language church worship was switched to English, the Sons of Norway decreed that English would be the language of their society, newspapers failed, and Norwegian summer schools closed. By the 1980s, Norwegian had contracted to the point where its sole domain was as family language, with a gap in transmission so that it has now almost completely disappeared (Haugen 1989: 62-69). Negative perception of minority languages has helped aid in the contraction of and shift from Breton in France as well. The language ideologies of France particularize Breton and characterize it in opposition to French: politically Breton is a regional language while French is the national language, Breton represented as the language of the past and of backward peasants while French is the

language of progress and the future, and Breton is a marker of local, rural identity while French is an international and urban language (Kuter 1989: 76). French *de jure* and *de facto* language policies have limited the domains of Breton, and attempted to make various public realms French-only, including the army, schools, and administrative offices, while all social and cultural institutions are dominated by French speakers (ibid. 80). In addition, rejection of the “old ways” along with the negative socio-economic Breton identity has included the abandonment of traditional music and dance forms. Except for self-consciously constructed revitalization projects, Breton has “little public presence” (ibid. 86) and its main functional domain is in the home, where, once again, transmission has been interrupted and Breton-speaking homes are often associated with self-conscious revitalization movements.

Related to the contraction of functional domains of a minority language is the concurrent reduction of its stylistic options, due in part to its “restriction in contexts of usage” (King 1989: 146). While many contracting languages retain a surprising degree of complexity even when in their last generation of speakers (cf e.g. Mithun 1989), stylistic shrinkage is common. For example, in Pipil (a Uto-Aztecan language of El Salvador) we find the loss of traditional oral literary devices such as paired couplets, along with the loss of original passives, necessary for complete discourse control, and in Ocuilteco (an Oto-Pamean language of Central Mexico) there has been a loss of the formulaic ritual language used for religious ceremonies and marriage petitions (Campbell and Muntzel 1989: 195). Young Dyirbal speakers do not command the traditional “mother-in-law” style of speech (used in the presence of taboo relatives), and almost all song styles have been lost as well (Schmidt 1991: 120). Terminal speakers of Arvanítika Albanian in Greece lack communicative competence in addition to grammatical competence: note that this is not

always the case, and Dorian (1981) found that semi-speakers of East Sutherland Gaelic often had an extremely high level of communicative competence that disguised their grammatical limitations. Terminal Albanian speakers not only make grammatical errors, but also accidentally use obscenities, depend on formulaic material, and use that formulaic material inappropriately; for example, in a recorded conversation a young man uses a formulaic expression that should be used only by older female speakers (Tsitsipis 1989). In Gros Ventre, an obsolescent Arapaho language of Montana (the last fully fluent speakers died in the 1980s), distinctive male and female phonologies are far less distinctive than in the past: an elderly male speaker uses female pronunciation, as do children in the bilingual program, since teachers and aides are all female (Taylor 1989: 173). And while a great deal of interspeaker variation remains in Newfoundland French (e.g., variation in agreement marking, in cliticization of object pronouns, and in number and case marking in the 3rd person), this variation “does not appear to carry the social meaning one finds in healthier speech communities” (King 1989: 146). Dressler (1982) suggested that the stylistic shrinkage associated with “language decay” could lead to monostylism in an endangered language, and Dorian (1994) set out to test this hypothesis using data from moribund East Sutherland Gaelic, a language with only domestic functional domains. She found that there were indeed markers of style shift; however, the range of styles was far more limited than is found in a language with both public and private functional domains, and one of the most salient stylistic markers was not taken from native Gaelic resources but was instead the use or non-use of English loanwords, where more “pure” Gaelic speech was associated with more formal narrative styles.

A third type of language attrition is described in the Matrix Language Turnover hypothesis devised by Myers-Scotton (1993, 1998 *inter alia*) to account for what is referred to elsewhere in the literature as structural borrowing. She proposes that languages “can sustain structural incursions and remain robust, but the taking in of alien inflections and function words is often a step leading to language attrition and language death” (1998: 289). Myers-Scotton’s hypothesis is that one mechanism of structural borrowing in language contact situations that have “bilingual speech” (usually code-switching, with morphemes and/or lexical structure from both of the languages in contact) is a “turnover” in the matrix language. A matrix language (defined and described at length in Myers-Scotton 1993) is a “morphosyntactic frame builder” in bilingual speech into which islands from the embedded language are inserted (Myers-Scotton 1998: 290). The Matrix Language Turnover has three phases. Phase One is characterized by frequent intrasentential code-switching with system morphemes from the matrix language and content morphemes from the embedded language, calquing that leads to utterances that would not be comprehensible to monolingual speakers of the matrix language, and the overgeneralization of matrix language categories such as tense to match their embedded language counterparts. In Phase Two the grammatical frame is a composite language based on both the original matrix language and embedded language. There is an increase in embedded language lexical items and calques, and the introduction of embedded language system morphemes such as participles and plurals, particularly in conjunction with embedded-language content morphemes. The lexical-conceptual and morphological realization structures come from both the matrix language and embedded language, and when there are several matrix language varieties of a structure, the one that corresponds to an embedded language variety will be preferred.

Finally, in Phase Three there is a matrix language turnover, where the new matrix language is the former embedded language. System morphemes are from the new matrix language and content morphemes can be from both languages, and the word order, lexical-conceptual structure, and predicate-argument structure are all from the new matrix language. The minority language that was once a primary language and is now a substrate is reduced to embedded content morphemes in the dominant language, content morphemes that may disappear altogether.

In the remainder of this chapter, I describe the theoretical and sociolinguistic background for the metalinguistic code-mixing found in Tatar and provide examples of this code-mixing and its associated structural innovations. In section 3.1.1 I describe the language attitudes and competence of the Tatars from whose speech this data was taken, and show that their code-mixing is in no way related to a lack of Tatar competence. In section 3.1.2 I define how ‘code-mixing,’ ‘discourse-pragmatics,’ and ‘metalinguistic’ are used in this chapter, and compare the pragmatic typology used here to classify code-mixed Russian words with typologies found in literature on discourse theory and pragmatics. Section 3.2 describes and briefly compares the varieties of Tatar performance produced by young bilinguals, as this metalinguistic code-mixing is found almost entirely in just one style of Tatar performance. In Section 3.3 I briefly present historical discourse-pragmatic borrowings that are found in present-day Tatar, and in Section 3.4 I present all of the code-mixed Russian words found in my Tatar corpus, along with illustrative examples taken from natural speech and analyses of morphosyntactic innovations associated with the code-mixing. Using discourse-pragmatic functionality as an organizing principle predicts that non-content words

and morphemes without this functionality will not be code-mixed; in section 3.5, I will present evidence showing that this is, indeed, the case.

3.1.1 Data Selection

The unconscious code-mixing of Russian words into Tatar and the language attrition that it signifies is particularly interesting because of the language competence and language ideologies of the speakers involved: they are the most competent and most ideologically devoted Tatar speakers of their generation to be found in urban Tatarstan. In Chapter One I described in detail the locus of my fieldwork: the “Tatar Social Club,” an informal and democratically run club founded in 1997 with the explicit goal of establishing a Tatar-friendly venue where Tatarphone youth, particularly those living away from their home villages, could meet and socialize with other Tatarphone youth. While the club is ostensibly politically neutral and does not accept outside funding of any kind, it is known to have a nationalistically oriented membership, including several former leaders of youth nationalism movements. Club members can often be found in prominent places at protests and rallies, and the meetings themselves take place on the second floor of a building housing a nationalist organization: the banner on top of this building, sitting on the roof just feet above the heads of club members and facing one of the main thoroughfares of the city, states: *Makhsatibiz: Tatarstanning Băysezlege* ‘Our Goal: Tatarstan’s Independence.’

Club members are considered to be among the city’s intellectual elite, and are described in articles and radio stories as “the cream of Tatar youth”: they are all either in college or graduate school, or have already received degrees and are working in one of the more highly regarded professions – for many of these professions (e.g., theater manager,

journalist, radio DJ) Tatar is used in the workplace. While attitudes towards the Tatar language vary widely among the populace as a whole, with Tatar generally given low prestige both by Russians and by many assimilated Tatars, particularly those under 30, among the Tatar intelligentsia and members of the Tatar Social Club the Tatar language – particularly *saf tatar tele* ‘pure Tatar’ – is awarded very high prestige. Use of Tatar is explicitly encouraged at club meetings, while Russian use is explicitly discouraged.¹

The data presented in this chapter come from yet a further subset of these Tatarphone youth. While club members range in proficiency from passively bilingual or semi-speakers (Dorian 1981) to fully competent and even Tatar-dominant, all of the speakers cited here are either Tatar-dominant or balanced bilinguals.² Tatar was the home language of their youth, and is either the language of education at the college level (e.g., in the Tatar Philology department of the university) or is currently the language of the workplace (e.g., a Tatar-language newspaper). They can therefore be seen as commanding Tatar in a variety of functional domains and registers. Additionally, these speakers are all actively engaged in Tatar culture and politics beyond participation in the Tatar Social Club. Moreover, several of them have repeatedly demonstrated that they choose ideology over communication: for example, refusing to switch to Russian to accommodate an uncomprehending audience, or addressing unknown people in Tatar rather than in Russian – in other words, they consciously violate the sociolinguistic conventions of Tatar use. Some of these speakers have gone so far as to explicitly state that they “never use Russian unless it is impossible to avoid.” Even so, these fully competent, culturally and politically aware, and ideologically

¹ I have seen people shout at a (balanced bilingual) club member making an announcement in code-switched Tatar and Russian, yelling “*Tatarcha!*” or “*Tatarcha gina!*” ‘Say it in Tatar!’, ‘Say it in Tatar only!’

² This classification is based both on my assessment and the assessments of their peers.

devoted speakers will produce these code-mixed Russian discourse-pragmatic words in certain situations when they believe themselves to be speaking “pure” Tatar.

3.1.2 Terminology

At this point, I would like to both clarify and define the terms *code-switching*, *code-mixing*, and *discourse-pragmatic words*, as they are all key to the analysis presented in this chapter.

3.1.2a Code-mixing vs. code-switching

Although some linguists use the terms “code-mixing” and “code-switching” interchangeably, I am following the distinction laid out by Olshtain and Blum-Kulka (1989). Code-switching is “when the speaker alternates units from different codes that are higher-level constituents, at least grammatical clauses or sentences,” while code-mixing “refers to smaller units, usually words or idiomatic expressions, which are borrowed from one language and inserted into the sentence of another language” (ibid. 60).

As will be seen in more detail in Chapter Four, the Hebrew words found to be code-mixed into English in Olshtain and Blum-Kulka’s corpus were mostly content words, in particular nouns and verbs. This is in contrast to my findings, where the code-mixing of Russian nouns and verbs, particularly where there is no lexical gap, was exceedingly rare.

3.1.2b Discourse-pragmatic words

The unconsciously code-mixed Russian words found in the speech and writing of the young Tatars described here belong to a variety of grammatical classes – including particles, adverbs, conjunctions, and question words – but can be grouped into the single functional

class of discourse-pragmatic words.³ It is notoriously difficult to find a conventional and cross-linguistic definition of just what comprises the class of discourse-pragmatic words. Much has been written on the subject of ‘discourse markers’ since Schiffrin’s groundbreaking 1987 book, but significantly less has been written on the category of discourse words in general – the Russian words code-mixed in Tatar-preferred style include, but are not limited to discourse markers. They can, however, all be interpreted as belonging to the set of what Fraser (1996, *inter alia*) calls ‘pragmatic markers,’ where discourse markers are just one subset of pragmatic markers; this pragmatic typology will be discussed at length below. I will begin this section with a discussion of discourse markers, as most scholarly work on discourse-pragmatics is devoted to this topic, and will then expand the discussion to include discourse-pragmatic words in general.

There is a bewildering array of terminology and definitions for the set of what I will be referring to as *discourse markers*; recent introductions to the topic usually note the most popular terms – *discourse marker*, *discourse particle*, *pragmatic particle*, *pragmatic expression*, and *connective*⁴ – and settle on *discourse marker* as the preferred term (cf. Jucker and Ziv 1998: 1, Schourup 1999: 229). However, there are significantly more than five terms used in the literature: Fraser (1998: 301) gives a list of fifteen alternative labels for what he refers to as

³ This is with the exception of a very small number of nouns, verbs, and adjectives, almost all of which denote culture-specific phenomena or fill lexical gaps in Tatar.

⁴ Representative uses of these terms are as follows. Schiffrin (1987 *passim*) uses the term *discourse marker*; Schourup (1985) and Abraham (1991) use the term *discourse particle*; Östman (1981) uses *pragmatic particle*; Erman (1987) uses *pragmatic expression*; and Blakemore (1987) uses *connective* (Jucker and Ziv 1998: 1).

discourse markers and Brinton (1996: 29) presents 23 alternative labels for what she calls *pragmatic markers*.⁵ In addition, there is no consensus on exactly which words and phrases constitute the class of discourse markers, both cross-linguistically and language-specifically, even for English, to which the vast majority of discourse studies are devoted (Jucker and Ziv 1988: 2), and Schourop (1999: 229) notes that the class of words referred to as *discourse particles* seems to be more inclusive than the class of words referred to as *discourse markers*. The functionality of discourse markers is also perceived differently according to the theoretical framework within which an analysis is based: discourse markers have been variously described as discourse connectors, turn-takers, confirmation-seekers, intimacy signals, topic-switchers, hesitation markers, boundary markers, fillers, prompters, repair markers, attitude markers and hedging devices (Jucker and Ziv 1988: 1). The functional domains of discourse markers are described as textual, attitudinal, cognitive, and interactional, and they are analyzed as text-structuring devices, modality or attitude indicators, markers of speaker-hearer intentions and relationships, and instructions on the processing of utterances (ibid. 4). Jucker and Ziv (1988: 2-3) suggest that discourse marker is a “fuzzy concept” with a “scalar conception of membership”: there is a range of features associated with discourse markers, and while none of the discourse markers have all of these features, there are Roschean “prototypical” members of the class of discourse markers – the more features a word has, the more it will be perceived as a prototypical discourse marker.

⁵ Fraser’s (1998) list is: *cue phrases, discourse connectives, discourse operators, discourse particles, discourse signaling devices, extrasentential links, indicating devices, phatic connectives, pragmatic connectives, pragmatic formatives, pragmatic markers, pragmatic particles, semantic conjuncts, and sentence connectives*. Brinton’s (1996) list is: *discourse marker, comment clause, connective, continuer, discourse connective, discourse-deictic item, discourse operator, discourse particle, discourse-shift marker, discourse word, filler, fumble, gambit, hedge, initiator, interjection, marker, marker of pragmatic structure, parenthetical phrase, (void) pragmatic connective, pragmatic expression, pragmatic particle, and reaction signal*.

Hölker (1991: 78-9, cited in Jucker and Ziv 1998: 3) suggests four characteristic

features of discourse markers:

- they do not affect the truth conditions of an utterance
- they do not add to the propositional content of an utterance
- they are related to the speech situation and not the situation that is being talked about
- they have emotive and expressive function rather than referential, denotative, or cognitive function.

Brinton (1996: 33-5) puts forth a different list of features, where discourse markers are:

- often in sentence-initial position
- outside of or loosely attached to syntactic structure
- optional in the sentence
- grammatically heterogeneous (including interjections, adverbs, particles, function words, verbs, conjunctions, phrases, idioms, and clauses)
- often phonologically reduced, forming a separate tone group
- (almost) without propositional meaning
- features of oral rather than written discourse
- able to appear with high frequency.

Schourup (1999: 230-233), synthesizing a decade of theoretical work, suggests that the main characteristics of discourse markers are their:

- connectivity, relating utterances to the propositions or implicatures of previous utterances, or to the context in a wider sense
- syntactic optionality
- non-truth-conditionality.

Using all of these characteristic features as diagnostic criteria, only five of the code-mixed Russian words found in Tatar-preferred style would be interpreted as prototypical discourse markers: *vot* 'there', *vot tak* 'so there', *nu* 'so, well', *no* 'however,' and *slushai* 'listen.' However, as will be seen in section 3.4, the set of Russian discourse-pragmatic words code-mixed in Tatar-preferred style is significantly larger; they are not all discourse markers *per se*, even though they all linguistically encode pragmatic functionality.

Many analyses of discourse-marking words and phrases rely on some sort of dichotomy that separates them from other sorts of words and phrases. These dichotomies include:

- lower-order truth-conditional speech acts vs. higher-order commentative speech acts (Grice 1989)
- propositional indicators vs. illocutionary force indicators (Searle 1969)
- propositional language vs. ‘everything else’ (Fraser 1996)
- conceptual meaning vs. procedural meaning (Sperber and Wilson 1986, Blakemore 1987)
- representational function vs. computational function (Rouchota 1998)
- host syntactic plane vs. disjunct syntactic plane (Espinal 1991)
- denotative vs. metalinguistic language (Maschler 1994)

Most theories of discourse functionality follow from Gricean pragmatics and Searle’s speech act theories. Writings on discourse markers (e.g. Rouchota 1998, Schourup 1999) often cite the work of Grice, particularly his dichotomy between “ground-floor” statements that are truth-conditional and “higher order speech acts” that comment on these statements, where “speakers may be at one and the same time engaged in performing speech acts at different but related levels” (Grice 1989: 362). Searle’s demarcation between propositional indicators and indicators of illocutionary force is essentially the same; “the illocutionary force indicator shows how the proposition is to be taken, or...what illocutionary force the utterance is to have; that is, what illocutionary act the speaker is performing in the utterance of the sentence (1969: 30).

Relevance theory, which has as its architects Sperber and Wilson (see in particular Sperber and Wilson 1986) builds on and supersedes Grice, and “within relevance theory, discourse connectives are, generally, seen as encoding procedural meaning which constrains the implicatures of an utterance” (Rouchota 1998: 109). Key to relevance theory is the dichotomy of conceptual and procedural meaning. Conceptual representations have concepts as their constituents (rather than syntactic categories), have logical properties, and

are usually part of the truth-conditional content of an utterance (Rouchota 1998: 111). Procedural meaning given by discourse markers (which are prime examples of procedural expressions) provides information on how to interpret the message of one utterance vis-à-vis the interpretation of a previous utterance (Fraser 1998: 302; Rouchota 1998: 111). In addition to Schiffrin's well-known formulation of discourse markers as "sequentially dependent elements which bracket units of talk" (1987: 31), she also suggests that "markers propose the contextual coordinates within which an utterance is produced and designed to be interpreted" (ibid. 315). Discourse markers have varying "contextual effects," which include strengthening an already existing assumption (*after all, furthermore*), contradicting and eliminating an already existing assumption (*but, however*), and combining with old information to yield new information (*therefore, then*) (Rouchota 1998: 112). For example, contrastive discourse markers in English have one of three procedural meanings: (1) they can contrast the explicit message of S2 [Statement 2] with the explicit or indirect message of S1 [Statement 1] (*but, however*); (2) they can show that the explicit message of S2 corrects the message of S1 (*instead of, rather*); or (3) they can show that S2, which is explicit, is correct, while S1 is false (*on the contrary*) (Fraser 1998: 306-307). In other words, procedural information can be understood "as a way of constraining or guiding the inferential phase of communication...utterance interpretation is largely an inferential process which consists in forming and evaluating hypotheses about what the speaker intends to communicate"; the theory is referred to as 'relevance theory' because procedural information "restricts the number of hypotheses the hearer needs to consider in order to arrive at an optimally relevant interpretation, thus facilitating his task" (Rouchota 1998: 111). The difference between

conceptual and procedural meaning can thus be reinterpreted as the cognitive distinction between representation and computation (ibid. 114).

Traugott (1997: 9), however, suggests that in studies of pragmatic implicature (particularly implicature that leads to grammaticalization), the focus should be not on the hearer who is processing discourse, but on the speaker who is producing it. In my analysis of the code-mixing of Russian discourse-pragmatic words, the focus will be on the production of speakers rather than on the processing of hearers, particularly with regards to the production of composite morphosyntax. Espinal (1991), writing on disjunct constituents such as adverbial phrases (*frankly*) and noun phrases (*Ladies and Gentlemen*),⁶ suggests that they reside on a different syntactic plane from the host sentence in which they are embedded. She is working within a syntactic “non-linear model” that is comparable to autosegmental phonology, but here the different levels are not assigned to phonological constituents such as consonants and vowels; rather, the syntax of the host sentence is on one plane, the syntax of the disjunct constituent is on the other plane, and the linear realization of the sentence is found at the intersection of the two planes where they meet. This syntactic model, designed for monolingual data, suggests a potential means of representing the mixed Tatar-Russian data presented in section 3.4. The Tatar host sentence can be analyzed as residing on one plane and the Russian discourse-pragmatic words found in the host sentence can be analyzed as residing on a different syntactic plane, one where they retain their Russian

⁶ What Espinal labels *disjunct constituents* are also referred to as *parenthetical expressions*, although the set of parenthetical expressions appears to be more limited. Espinal includes the following items in the class of disjunct constituents: disjunct sentences, appositive relatives, adjectival phrases, adverbial clauses, noun phrases, prepositional phrases, combinations of any of the above, speaker-oriented adverbs, vocatives, and discourse adverbials.

morphosyntactic requirements; the linear intersection of these two syntactic planes produces a composite morphosyntax when the Tatar and Russian morphosyntax are not congruent.

The two remaining dichotomies are Fraser's (1996) separation of propositional language and 'everything else' and Maschler's (1994) separation of denotative and metalinguistic language. These two analyses of language are complementary, and I will use both to classify the Russian non-content words code-mixed in Tatar-preferred style. Maschler coined the term "metalinguaging" to describe the non-content Hebrew words (mostly discourse markers) that she found code-mixed in the English speech of English-Hebrew bilinguals in Israel. Essential to the classification is the separation of discourse into two levels: the *denotative* level, which is oriented to the world, and the *metalinguistic* level, which is oriented to the process of using language (1994: 327). It is the words used in this second level of discourse, the metalinguistic, that Maschler labels *metalinguaging words* and that I will be referring to as *discourse-pragmatic words*. Maschler uses *metalinguaging* rather than *metalinguistic* because she is highlighting that language is a dynamic process rather than a static object; I too regard the phenomena presented here as processual, and at the intersection of language as a practice and language as a system. Although many metalinguaging words can be interpreted as procedural or computational according to relevance theory, the focus here is on speaker production and not hearer interpretation. However, future research on bilingual code-mixing and style-shifting should perhaps include cognitive aspects of both speaking and interpreting; hence, I prefer the more neutral term *discourse-pragmatic words*, as it does not focus exclusively on either side of the dialogue. Fraser's (1996) dichotomy of propositional and pragmatic is similar to the dichotomy of denotative and metalinguistic, but placed within a different theoretical framework; he too is focused on the speaker. In this

framework the non-propositional elements of a sentence can be analyzed as pragmatic markers that are “the linguistically encoded clues which signal the speaker’s potential communicative intentions” (1996: 168).

In Fraser’s pragmatic typology (cf. Fraser 1988, 1990, 1996, 1998), which was designed to account for monolingual English pragmatic phenomena, discourse markers are seen as a subclass of pragmatic markers. There are four types of pragmatic markers, which relate to four message types (Fraser 1996):

(1) Basic markers. These pragmatic markers specify the force and the nature of the propositional message, and can be lexical, structural, and a hybrid of the two. Lexical basic markers include performatives such as *I promise*, and also include pragmatic idioms like *please*. Structural basic markers are based on syntactic mood: declarative, imperative, or interrogative. Hybrid basic markers combine lexical items and syntax, and include tag questions.

(2) Commentary markers. These pragmatic markers comment on the basic propositional message, and have both representational and procedural meaning. They include markers of assessment (*sadly*), manner of speaking (*frankly*), emphasis (*mark my words*), mitigation (*if you don’t mind*), consequent-effect (*to sum up*), evidential markers (*certainly*), and hearsay markers (*allegedly*).

(3) Parallel markers. These markers give a message that is parallel and in addition to the basic message, e.g. ‘*I regret that he is still here,*’ and include vocatives.

(4) Discourse markers. Discourse markers relate the basic message to the preceding discourse. There are four types of discourse markers in Fraser’s typology (1996: 187-188):

- Topic change markers (e.g., English *in fact, now*; in my data, Russian *slushbai* ‘listen!’)
- Contrastive markers (e.g., English *but, nevertheless*; in my data, Russian *no* ‘however’)
- Elaboration markers (e.g., English *above all*; in my data, Russian *to est* ‘that is’)
- Inferential markers (e.g., English *after all, as a result*; in my data, Russian *tak, chto* ‘so’)

While Fraser’s system has been criticized for being overly inclusive (Schourup 1999: 239), this pragmatic typology aptly captures the pragmatic functions that mark the Russian code-mixed words found in Tatar as a single functional class of discourse-pragmatic words.

As will be seen in section 3.4 below, I have grouped the code-mixed Russian metalinguistic words found in my Tatar corpus into three major categories: (1) Markers of discourse structure and force; (2) Interactional performatives, and (3) Evaluatives.⁷ These categories do not follow Fraser’s organizational pattern, but all of the words found therein can be classified according to his pragmatic typology. The set of words that I am classifying as markers of discourse structure and force include, but are not limited to, what Fraser calls discourse markers. In addition to functioning as markers of topic change, contrast, elaboration, and inference, they more generally create discourse and narrative structure and comment upon discourse; they are all procedural connectives or commentaries. My second set of discourse-pragmatic pragmatic words, interactional performatives, is congruent with Fraser’s lexical basic markers, as they are all pragmatic idioms. Finally, the set of words that I am classifying as evaluatives includes words that Fraser would classify as lexical basic markers, commentary markers, and parallel markers. For example, what I call discourse-pragmatic words of positive evaluation are equivalent to Fraser’s lexical basic markers in the category of idioms – what he refers to as idioms that express pleasure (class c) and idioms that express agreement (class e) (1996: 180). Some of these Russian evaluative words,

⁷ I am most grateful to Eve Sweetser for her help with assessing the discourse-pragmatic nature of these words and with their organization into these three categories and related sub-categories.

particularly those expressing agreement or evaluating applicability, would be classified as discourse markers in other theoretical frameworks; however, what they all have in common is their expression of speaker stance and evaluation of discourse, therefore I have chosen to group them together.

In Fraser's pragmatic framework, the first three sets of pragmatic markers (basic markers, commentary markers, and parallel markers) are seen as propositional and containing representational meaning, while discourse markers have procedural meaning "and specify how the sentence of which they are a part is related to the preceding discourse" (1996: 182). In other words, discourse-pragmatic functionality is found not only in words that have procedural meaning, but also in words with propositional meaning. Integrating Fraser's and Maschler's frameworks shows that while all denotative language is propositional (as it references the world), metalinguistic language too can be propositional. All of the Russian code-mixed words found in Tatar are metalinguistic in function, that is, they refer to and organize language, and include both procedural and propositional words.

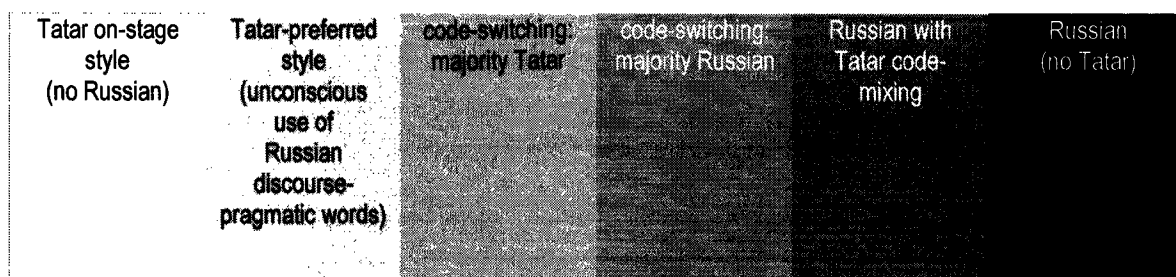
Propositional language is one level of language in the functional-semantic model found in Traugott (1982), which is based on the model proposed by Halliday and Hasan (1976). Traugott's model has three functional-semantic components (1982: 247-248). The first is the propositional component, which involves the resources of the language that "mak[e] it possible to talk about something." This is the main locus of truth-conditional relations, but also includes deictics to places (*here-there*), times (*now-then*), and persons (*I-you*). The second is the textual component, which "involves the resources available for creating a cohesive discourse" and includes connectives (*but, therefore*), anaphoric and cataphoric pronouns, topicalizers, relativizers, and complementizers, which all share the property of

being directly linked to the unfolding of the speech event itself. The third component is the expressive component, which involves “the resources a language has for expressing personal attitudes to what is being talked about, to the text itself, and to others in the speech situation.” Expressive words and phrases show cohesion, attitudes towards propositions that cohere, and evaluations of propositions that cohere, and also include honorifics and markers of turn taking. The process of grammaticalization, which is unidirectional, moves from less to more personal in each component, and takes the propositional level of language as the starting point. Grammaticalized words either move from being propositional to textual (and then perhaps are grammaticalized again so that they are expressive) or move directly from the propositional level to the expressive level of language (Traugott 1982, 1989). In Traugott’s system, while words in all functional-semantic levels can have pragmatic functionality, metalinguistic words tend to be words that have undergone grammaticalization. Additionally, Dancygier (1992) suggests differentiating between metatextual and metalinguistic words, where metatextual words comment on the interpretation of the text and metalinguistic words comment on the form of the text. However, the words that I am categorizing as discourse-pragmatic and metalinguistic terms come from all three functional-semantic levels of Traugott’s model, and the data do not indicate a necessary separation into metatextual and metalinguistic. In conclusion, the Russian non-content words code-mixed into Tatar are metalinguistic but include words that are not procedural in function; they can be propositional, procedural, or both; they include metatextual words; and while they all have discourse-pragmatic functionality, they are not limited to prototypical discourse markers.

3.2. Varieties of Tatar performance

Integral to understanding the phenomenon of unconscious code-mixing of dominant-language words is an understanding of the linguistic performance and style shifting of young Tatar-Russian bilinguals.⁸ Although as language investigator my presence would trigger exceptionally high levels of language awareness and verbal hygiene, thus limiting my access to the style shifting of club members, after months of observation I discovered that urban bilingual Tatars had a continuum of linguistic performance that was related to language mixing, with pure Russian on one end and pure Tatar on the other (this was briefly described in Chapter One, section 1.3.1). Figure 13 is a graphical representation of these styles.

Figure 13. Cline of language mixing and performance styles for urban Tatar bilinguals



For the sake of clarity, these styles are arranged according to a single stylistic variable, the level of Russification at the lexical level. I should note that this representation is overly simplistic in two ways. The first is that there are other markers that can differentiate one style from another, although word choice is the most salient and also most relevant for the findings described in this chapter. The second is that each style, represented here as a shaded

box, is not a monolithic entity, but rather a collection of heterogeneous types of linguistic performance. For example, Tatar on-stage style can be in a variety of registers: formal and informal, to name just two. Additionally, there are nuances that differentiate different kinds of performance within these registers, even when they are in the same channel or setting; for example, the speech of a master of ceremonies at a Tatar concert will be different from the speech of a master of ceremonies at a literary tribute evening, although both will be in Tatar on-stage style, with the very highest level of verbal hygiene, and without any Russian words that are not conventional borrowings. ‘Style’ and ‘register’ are both slippery terms and used variously by linguists, although ‘style’ is most often used to describe level of formality or placement on a cline of status vs. solidarity, while register is usually related to the type of language used in a particular context. Romaine (1994: 20-21) gives fairly representative definitions, and characterizes register as involving “consideration of the situation or context of use, the purpose, subject matter, and content of the message, and the relationship between participants,” while style is related to register and is dependent on the social context, the physical environment, the topic, the relationship of the participants, and personal characteristics such as their social class and age. If one follows these definitions as guidelines, one main way to distinguish among registers, also referred to as ‘genres’, is through vocabulary differences; examples of registers include literary language, political language, medical language, and legalese (Verschueren 1999: 118). However, I am following Tannen (1984) in giving ‘style’ a more broadly inclusive definition: style is “the mix

⁸ I have observed this style shifting in both urban and rural Tatars up to the age of about 50. It seems highly likely that the linguistic performance and style shifting of elderly rural Tatars is different from what is described here; however, their linguistic performance is outside the scope of this dissertation.

of devices speakers use in different contexts” and refers “to no more nor less than a way of doing something.” When describing the speech of young Tatar bilinguals, I am using the term ‘style’ to represent a set of clearly related types of linguistic performance – a set of performance types that is in opposition to other sets of performance types, although it is possible to have overlap at the boundaries of these sets. Therefore, although the term ‘style shifting’ has been used elsewhere to describe what I am referring to here as ‘code-switching,’ which is to say, switching between two languages (or in some cases, between two dialects), I differentiate between style shifting and code-switching; in my proposed continuum of Tatar linguistic performance, code-switching is a style in and of itself.^{9,10} In differentiating between code-switching where the majority of discourse is in Russian and code-switching where the majority of discourse is in Tatar, I am following Myers-Scotton’s conception of code-switching (Myers-Scotton 1993, 1998), where what is central is which language is being used as the matrix language and which language is being used as the embedding language. This distinction is more relevant for this performance continuum than the metaphorical vs. situational code-switching distinction devised by Blom and Gumperz (1972). However, I am leery of the decision-making process involved in discerning which language in a stretch of discourse is the matrix language and which is the embedding language – this is difficult in all code-switching situations, and Tatar-Russian contact is no exception, particularly given the sometimes playful adaptation of Russian words to Tatar phonotactics that periodically appears in speech and writing, e.g., *maladis* for Russian *molodits* ‘well done’. Moreover, the matrix language does not need to be consistent for an entire stretch of discourse, and can

⁹ Actually, code-switching is two styles, depending upon which language is the majority language.

¹⁰ As used here, a code is “any distinguishable variant of a language, involving systematic sets of choices, whether linked to a specific geographical area, a social class, an assignment of functions, or a specific context of use” (Verschuere 1999: 118).

switch from sentence to sentence. Therefore, I will use the more general term ‘majority language’ rather than ‘matrix language’ as it makes no claims as to the morphosyntactic structure of the code-mixed discourse, but highlights the most salient aspect of that discourse – that it is either predominantly Tatar or predominantly Russian.

3.2.1 “Tatar on-stage” style

“Tatar on-stage” style is the style with the very highest level of verbal hygiene (Cameron 1995) and self-consciousness, used most prominently in the construction of a cultural and ethnic Tatar identity. This identity construction can be found in performance for a variety of audiences: outsiders who are non-Tatars; people who are simultaneously in-group and out-group, which is to say ethnically Tatar but personally not known to the speaker; and insiders, members of a local and “dense” social network (Milroy 1987). In other words, when performing in this style – and it is very much a “performed” style – Tatars are performing both for others and for themselves.

Tatar on-stage style is the realization of speakers’ attempts to produce “pure” and/or “literary” Tatar (known locally as *saf tatar tele* and *ädäbi tatar tele* respectively), and is most noticeably characterized by the absence of Russian words. This de-Russification of speech can be interpreted as symbolic of the de-Russification of both Tatar culture and Tatar ethnic identity, which in post-Soviet Tatarstan are defined in opposition to Russian. This indexicality, where language, culture, and religion are inter-related to the point where invoking one of these aspects of Tatar identity automatically references the other two, was examined more fully in Chapter Two.

Tatar on-stage style is often found in the public sphere in formal registers when the setting, often literally on-stage, requires both a high literary standard and a presentation of Tatarness. This style is found most commonly in performances (in the most standard sense of the word) at Tatar cultural events such as lectures, tribute evenings, and concerts – the speech of both the master of ceremonies and the musicians will be in Tatar on-stage style, with the musicians usually speaking less formally. Tatar on-stage style can also be found on radio and television: in the speech of DJs; in speech by both parties in an interview; in all speech produced by newscasters, be it scripted or off-the-cuff; and by radio listeners who call in to make requests or comments. The end result is that there is no Russian to be heard in present-day Tatar media, with the exception of conventional borrowings. Political speeches, particularly those speeches televised from the floor of the republic’s parliament, are also in Tatar on-stage style. Off-stage, Tatar on-stage style can be found in announcements, statements, and comments addressed to the general crowd at meetings of the Tatar Social Club, or at other similar cultural events. Finally, conversations with or for investigators of Tatar language and culture will be performed in Tatar on-stage style.¹¹ (The effect of the fieldworker on linguistic performance was discussed more fully in Chapter One.)

Even Russian-dominant speakers, who are not fully competent in Tatar, will attempt to completely de-Russify their speech when it is called for by either setting or audience, which can lead to ungrammatical, and sometimes incoherent, speech. However, the

¹¹ An anecdotal example: during the intermission of a concert by my favorite Tatar singer, a friend took me backstage to meet him. He had been bantering with the audience in Tatar only, and when conversing with me (self-identified as an American and Tatar language learner) did the same: however, he would periodically turn his head and converse with his band members in code-switched Tatar and Russian. Turning back to me would mark his return to Tatar on-stage style (or here, Tatar back-stage style).

requirements of culturally appropriate linguistic performance supersede those of fluid and articulate communication. This performance pressure can prove to be overwhelming for some speakers who are closer to the less-competent end of the cline: they may precede any statements with profuse apologies for their lack of Tatar competence, burst into tears, or flee rather than say anything at all.

3.2.2 “Tatar-preferred” style

By contrast, Tatar-preferred style is found mostly in private conversations where it has been established that Tatar is the preferred language of communication, usually due to audience and setting. This style is found mostly in informal register, and is more “unguarded”: the level of language awareness and verbal hygiene is lower than in Tatar on-stage style. Typical situations where Tatar-preferred style can be found include private conversations at the Tatar Social Club; a visit to the home of a peer whose parents are Tatar teachers (and present); and intergenerational family communication, particularly in the home setting.

The code-mixing of Russian discourse-pragmatic words is found for the most part in Tatar-preferred style, but can also be found in Tatar on-stage style (which is the style of the majority of the conversational recordings in my Tatar corpus). However, in Tatar on-stage style tokens are few and far between: when the filtering goes on for real, this code-mixing disappears.

3.2.3 Other Tatar styles

The remaining styles of Tatar performance are pure Russian with no Tatar lexical items, Russian with Tatar code-mixing, and the two code-switching styles (one where the majority of the discourse is in Tatar and the other where the majority of the discourse is in Russian).

As described in Chapter One, pure Russian will be used with monolingual Russian speakers and for public transactions where sociolinguistic conventions require that Russian be the language of conversation. The role of audience design in the specification of language is quite clear: speakers would often maintain a side conversation with me in Tatar while they were conducting their business in Russian; transacting business in Russian did not dictate a switch in the style of the private conversation.

Although in Figure 13 the style of Russian with Tatar code-mixing can be interpreted as the mirror image of Tatar-preferred style, it is actually quite different in both linguistic structure and language awareness. Tatar speakers appear to be unaware of the code-mixing of Russian words in Tatar-preferred style, and the code-mixed words are not content words but rather discourse-pragmatic words. By contrast, code-mixing of Tatar words into otherwise Russian discourse is quite deliberate and involves only content words or vocatives. Its purpose is as a marker of Tatar ethnic identity, and its goal is either a shift to one of the styles located on the more Tatar end of the cline, or an expression of solidarity because that solidarity is thought to be useful. Several young Tatars mentioned this Tatar code-mixing as a response to the interview question: “You speak Tatar with your friends and acquaintances. When you were becoming acquainted, how did you know that you could speak Tatar with

them? Could you talk about this process?”¹² I was able to observe this deliberate code-mixing firsthand when I injured my leg on a trip organized by the Tatar Social Club and was accompanied to a Kazan emergency room by several club members who had been on the trip. They seemed to feel that the service I was receiving in the X-ray department was substandard, and played the Tatar solidarity card by code-mixing Tatar into their otherwise Russian speech with the nurse in charge, calling her *Apa* ‘Aunty’, a respectful term of address, and using Tatar nouns for various objects in the scene. Unfortunately, this did not result in any sort of improvement in service, but they had clearly made an effort on my behalf, and when questioned later, admitted that the code-mixing had indeed been consciously done. I was also informed that this style was reputed to be useful if stopped by a police officer on traffic duty who had a Tatar appearance, and that clever use of Tatar could allow one to avoid a ticket – however, as none of my informants claimed first-hand knowledge in this matter, it is difficult to determine if this is fact or merely wishful thinking.

Unfortunately, I am unable at this time to give a detailed and precise description of the two Tatar-Russian code-switched styles. As mentioned in both Chapter One and earlier in this section, the language ideologies and linguistic construction of identity of the young Tatars with whom I worked did not allow them to code-switch in my presence – the requirement to speak “good” Tatar in front of a language investigator meant that for the first six months or so of my fieldwork I was presented only with Tatar on-stage style, and only later in my fieldwork did some speakers relax their verbal hygiene sufficiently to use Tatar-preferred style with or near me (other speakers never used anything but Tatar on-stage style in my presence for the duration of my fieldwork). I was only able to periodically overhear

¹² In Tatar, *Sin khäzär duslarıng belän yaki tanışlarıng belän tatarcha söyläshäsäng. Alar belän tanışkanda, sin alarıng tatarcha belüleren nichek beldeng? Şul protsess turında söylächi.*

code-switching, most often by Tatars who were not members of the Tatar Social Club or simply not known to me, and was able to record general impressions of the lexical items and grammatical structure involved after the fact: none of the recordings of Tatar Social Club members, even those recorded with “home-recording kits” in my absence, produced any examples of code-switching. This reluctance to code-switch may not be entirely dependent on the presence of the language investigator: while some young Tatars will code-switch in their postings on internet bulletin boards, the subset of Tatar Social Club members whose spoken and written data are presented in section 3.4 will not code-switch in their written discourse on the internet, and will write in the equivalents of Tatar on-stage style and Tatar-preferred style.

The code-switching constraints, triggers, and boundaries for the Tatar speakers whose discourse is presented here can therefore only be analyzed at a later date, after this style-shifting obstacle has been surmounted. I can, however, offer one example of metacommentary on code-switching that gives some insight into both the triggers for switching and the attitudes of some young Tatars towards code-switching styles. One evening, at around 3 am, I was sitting around a campfire with approximately six members of the Tatar Social Club and about five more of their friends (this was the same trip mentioned above, and fatigue induced by this campfire conversation can perhaps be blamed for the leg injury that occurred just a few hours later). All of the Tatar Social Club members in the conversation were students in the same department, and training to be philologists, and they asked me a series of questions in quick succession about the structure of English. I laughed and said the following:

[3.1]
Yarar, min khazär *leksiia* **chita-iu**
Ok, I now *lecture* **read-PRES.1ps**
'Ok, now I'm **giving** you a *lecture*.'¹³

One of the young women laughed in response and said something along the lines of “Now you’re talking like one of us!” The word *leksiia* ‘lecture’ is a conventional borrowing from Russian that requires the verb ‘to read’ in both Russian and Tatar (*chitat’* in Russian, *ukirga* in Tatar); its Russian provenance appears to have triggered my shift into Russian, such that I ended the sentence verb-finally, as one would in Tatar, but with an appropriately conjugated Russian verb. (Note that this sentence produced entirely in Russian could end verb-finally, but with the effect of emphasizing the verb.) The young woman’s response suggests that conventional Russian borrowings act as triggers for code-switching into Russian, and her positive assessment of my code-switching, that it made me more like “one of us,” implies both that young Tatars code-switch and that this code-switching is viewed positively and as a marker of youth speech.

3.2.4 Methods of data collection

The style-shifting and linguistic performance of the Tatars among whom I was a participant observer were directly affected by my presence or by the presence of my recording equipment, both of which often triggered such a high level of verbal hygiene and self-conscious performance that unconscious code-mixing did not take place. In order to obtain

¹³ Transliteration conventions for Tatar and Russian can be found in Appendix A. The conventions for data presentation are as follows: Tatar is in plain text; while inserted Russian words are in **boldface** both in the data and in the translation. For the purposes of clarity, standard Russian borrowings are presented as Tatar. Subordinate clauses and periphrastic verbal constructions are periodically underlined for the sake of clarity, in order to show their syntactic relatedness.

the widest range of data and examples of code-mixed Russian words, I used four methods of data collection: recordings of conversations; fieldnotes; e-mail; and an internet bulletin board. Recordings were made in three ways: in my apartment while I was in another room; in my apartment while I was present; and using “home recording kits” that I sent home with various speakers (each kit included a tape recorder, two blank tapes, instructions, and letters of permission). Due to the requirements of the University of California, Berkeley’s Committee for the Protection of Human Subjects, both oral and written permission were received before beginning a recording. This level of attention to the recording process, in addition to the other “unnatural” factors of audience design inherent in the recording of a private speech event, meant that almost all of the recorded conversations were in Tatar on-stage style rather than any of the other styles one might expect to find in a “casual” interaction among peers. Even so, instances of code-mixed Russian words do appear in these recorded conversations.

Tokens are more frequent in data that I recorded in my fieldnotes, and in e-mails and internet bulletin board postings written by club members. Data that come from fieldnotes were written down in situations where I was either a conversation participant or a ratified auditor: sometimes these notes consist of only the word in question with commentary on how often it was used in the conversation, while at other times I was able to jot down the entire sentence in question. All of the conversations where code-mixing was noted and recorded in fieldnotes were in Tatar on-stage style and Tatar-preferred style, most often the latter.

Finally, we have the two written channels, both of which are informal and electronic, and appear to loosely correspond with Tatar-preferred style as spoken. The first is e-mail

that was sent to me by club members. All of this e-mail was necessarily written in one of Kazan's few internet salons, where computer time is purchased by the hour, and one hour costs approximately 30 rubles (about \$1 in the years 2000 and 2001). This sum represents a significant percentage of the salary of many young Tatars, particularly those who are students and receive a monthly stipend of approximately \$7 per *month*, a stipend that is meant to cover all living expenses in Kazan. (Even teachers and journalists earn only about \$50 a month.) Based on this fact, along with observation of young Tatars in internet salons, I believe that these e-mails are not written slowly, composed in advance, or edited for content. Time is of the essence, and they are dashed off and sent immediately after they are written. The same is true of postings on the bulletin board of a Tatar youth "get acquainted" website. The website itself is international in scope and refers to itself as "*Tatarlar öchen berenche tanishular saytı*" "the first site for introducing Tatars." The site contains various pages: personal ads of both a friendly and romantic nature, a guest book, a "brief course in the Tatar language," virtual postcards with headings in Tatar, and a "forum," which is the bulletin board in question. While the rest of the site is trilingual in Russian, Tatar, and English, with Russian dominating, the bulletin board has been established as a place where Tatar is the preferred medium of communication. Posters who write in Russian will often preface their posting with an apology for their lack of Tatar knowledge, and other posters will sometimes reprimand people for writing in Russian, further establishing Tatar as the standard that they are attempting to maintain. Although personal ads and guest book entries are written by Tatars in a variety of locations (including America, Japan, Turkey, and Finland), the majority of bulletin board posters for this website are located in Kazan, and are in fact members of the Tatar Social Club. Most of them post pseudonymously, but their pseudonyms are known

to me (and were usually checked in the most direct of ways – I simply asked them). Of the bulletin board postings written in Tatar, the vast majority are code-switched or have code-mixing of Russian discourse-pragmatic words. However, for the purposes of this dissertation, I used only material that was written by club members personally known to me, who met all the criteria listed in section 3.1 above, and for whom spoken data had already been collected.

3.2.5 Final thoughts on “purity”

Style shifting is set within the backdrop of a ubiquitous “discourse of purity” (cf. Hill 1998, Kroskrity 1998) that is found not only among these young Tatarphones, but is presented in newspapers and magazines, on television and the radio, and in private conversations. “Pure” Tatar (*saf tatar tele*) and literary Tatar (*ädabi tatar tele*) are explicit standards, and were examined in detail in Chapter Two. “Pure” Tatar, according to the various definitions given by Tatars, appears to be perceived as the following: Tatar, spoken in any domain and in any register, without any salient Russian interference. This pure Tatar is contrasted with “impure” Tatar, described variously as Russian phonetic interference, spelling mistakes, mistranslations, calques, and code-switching with Russian. Verbal hygiene can then be interpreted as, to a great extent, the de-Russification of Tatar. When the level of verbal hygiene is relaxed, as in Tatar-preferred mode, code-mixed discourse-pragmatic words (as if they are somehow “flying under the radar” of verbal hygiene) slip through, and Tatar speakers appear to be

unaware of them. In example [3.2] below, we see the response of E.¹⁴ when asked the question, “What do you think speaking “pure” Tatar (*saf tatarcha*) means?”

[3.2] What is “pure” Tatar?

Saf	tatarcha –	ul	rus	süz-lär-e-n	kistir-miycha
pure	Tatar –	it	Russian	word-PL-POSS-ACC	insert- NEG.PRES.GER
söyläshü,	tinglau-chi-lar-ga	priyatno	bul-irga	tiesh	tingla-rga.
conversation	listening-AGT-PL-DAT	pleasant	be-INF	should	listen-INF.

‘Pure Tatar – it is conversation without inserting Russian words, it should be **pleasant** for listeners to hear.’

This code-mixing of the evaluative Russian adverb *priyatno* speaks for itself.

3.3. Past function-word borrowing in Tatar

Tatar borrowing of non-content words, in particular function words, is not limited to the 20th and 21st centuries, with their increase in intensity of Russian contact and language shift. In fact, it has been claimed that “[t]he syntactic structure of Turkic has proved relatively open to foreign influence” and that most Turkic languages “have copied foreign relators, e.g., conjunctions, and sometimes also clause types typical of such elements” (Johanson 1998: 118). Like many Turkic languages, Tatar has historic superstrates of Persian and Arabic, which date from the 10th century Bolgar conversion to Islam. As superstrates,

¹⁴ Because some of this data comes from public domain material on a website, I am using initials to refer to the Tatar speakers quoted here, although all of them have pseudonyms that they chose themselves that are used in my other work. This is because some of these speakers do not post pseudonymously, or their screen names are known to others, and I want to ensure that there is no link possible between the identities given in this dissertation and real-world identities.

Persian and Arabic had a significant influence on literary written Tatar, with less of an influence on the vernacular.

The vast majority of the non-content words borrowed from Persian and Arabic, like their present-day Russian code-mixed counterparts, were metalinguistic discourse-pragmatic words used to structure and comment upon discourse. These words will be examined in depth in Chapter Four. Several of the borrowed Arabic and Persian discourse-pragmatic words were syntax-altering, resulting in changed word order, and in some cases, in changed clausal structure – changes that have perhaps made present-day Tatar more receptive to those Russian conjunctions that are associated with a similar syntax.

The two early grammatical borrowings from Russian are also discourse-pragmatic words:

- *ä* ‘and/but’
- *bit* discourse particle meaning ‘is it not?’, also emphatic particle

The conjunction *ä* ‘and/but’ comes from Russian *a* ‘and/but’, while the emphatic particle *bit* is Russian *ved’*, which is both a discourse particle and an emphatic particle. The phonetic nativization shows that these are early borrowings, as later borrowings remain true to Russian phonotactics. Russian *a* is a contrastive conjunction used to create discourse structure by opening a new topical unit – it “encodes a switch in the thematic dimension” (Grenoble 1998: 180). The particle *ved’* in Russian “references background knowledge which the speaker assumes that the interlocutor shares. Use of this particle not only activates this knowledge in the discourse, but implies its indisputability” (ibid. 24).

Tatar is not the only Turkic language to have borrowed Russian *a*: in [3.3] below, we see an example of Tuvan borrowing of *a* (Tuvan is a Siberian Turkic language of the Sayan

branch, spoken in regions of Russia not physically adjacent to those where Siberian Tatars can be found).

[3.3] Tuvan use of Russian *a*

oon chogum ad -i sendi, a shola-zī odaar
 he-GEN real name-POSS Sendi, **and/but** nickname-POSS Odaar.
 ‘His real name is Sendi, **and** his nickname Odaar’
 (Monguš and Sat 1969: 200, cited in Comrie 1981:35)

In example [3.4], written by a 21-year old journalist who was born and raised in Kazan, we can see that Tatar use of *ä* is parallel – here it is used to begin what is punctuated as a new sentence but could also be interpreted as a subordinate clause. The example also includes a bonus use of *bit*.

[3.4] Tatar use of *ä* and *bit*

Bu üz-ebez-neng tarikhi chiganak-lar-ni barla-u, boringi
 this self-1PP-GEN historical source-PL-ACC verify-V.NOUN ancient
 baba-lar-ibiz jir-en-ä säjdä kil-ü, alar-ning
 grandfather-PL-1PP earth-POSS-DAT bow do-V.NOUN they-GEN
 rukh-in-a bagishla-p doga uk-u da bit.
 soul-POSS-DAT devote-2nd GER prayer read-V.NOUN also **EMPH.PART**
 Ä shushü niyät-lär kem-gä-der osha-miy,
And/but such goal-PL who-DAT-INDEF.PART please-NEG.PRES.3PS
 örket-ä bugay.
 surprise-PRES.3PS it seems.

‘This is the verification our own historical origins, a journey made in order to bow down [in respect] to the land of our ancient grandfathers, and also the reading of a prayer devoted to their souls, **is it not? And (/but)** such goals displease someone, surprise them, it seems.’
 (Khäyrullina 2001)

Tatar has a long history of borrowing discourse-pragmatic function words; most of these discourse-pragmatic borrowings will be presented and analyzed in section 4.3 of Chapter Four. Additionally, the altered syntax that comes with the borrowing of some of these function words (e.g., a change from only embedded subordinate clauses to the use of subordinate conjunctions and clauses that follow the main clause) may in some way aid the present-day receptiveness to Russian discourse-pragmatic words that fill the same syntactic slots.

3.4. Present-day code-mixing of Russian words

When speaking in Tatar-preferred style and unconsciously code-mixing, young Tatars will code-mix very few verbs and nouns – those verbs and nouns that are inserted are usually culture-specific or fill a lexical gap. Once we remove verbs and nouns from consideration, the remaining code-mixed Russian words are seen to be metalinguistic words – they are all words with discourse-pragmatic function. Many of the code-mixed words are Russian discourse deictics, which fall into two general functional categories: (1) marking topical boundaries in text content, specifically the opening and closing of topics; and (2) linking elements of the text to the textual setting, including speakers' background and inferential knowledge (Grenoble 1998: 26). Other Russian discourse-pragmatic words are used for narrative structure, to express speaker evaluation and stance, and as interactional performatives that are themselves social gestures. If we use discourse-pragmatic functionality as the organizing principle for these code-mixed words, two predictions follow: (1) polysemous Russian words will be used only in their evaluative or grammaticalized meanings, that is, in those meanings that have discursive or pragmatic force, and (2) non-

content words and morphemes that do not have discourse-pragmatic function will not be code-mixed. These predictions in fact are seen to be true, and supporting evidence for this claim will be presented throughout this section as well as in section 3.5 below.

The code-mixed Russian discourse-pragmatic words presented below are organized into three main pragmatic groups: (1) Markers of discourse structure and force, (2) Interactional performatives, and (3) Evaluatives. The relationship between these groups and pragmatic and discourse typologies found elsewhere was discussed in section 3.1.2.

3.4.1 Markers of discourse structure and force

3.4.1a Discourse structure marking – subordinative

The first three code-mixed Russian words in this category are discourse-structuring subordinating conjunctions:

[3.5]

- *chto* ‘that’
- *chto(b)(y)* ‘in order to, that’
- *potomu chto* ‘because’

In [3.6] below, we see an example of an utterance with a code-mixed *chto* used as the subordinator – the subordinate clause is underlined.

[3.6] Code-mixed *chto*

Min shat singa, chto sin taza-sau häm kör küngele.
I happy you.DAT that you healthy and cheerful mood-with.
‘I’m happy for you that you’re healthy and in a cheerful mood.’ –A.

The Russian equivalent is given in example [3.7], with the subordinate clause also underlined.

[3.7] Russian equivalent

Ia rad dliia tebia, chto ty zdorova i schastliiva.
I happy for you-GEN, that you healthy and happy.
'I'm happy for you that you're healthy and cheerful/happy.'

A comparison of the glosses in [3.6] and [3.7] reveals that the Tatar utterance using a code-mixed *chto* 'that' is almost identical syntactically to Russian. The only difference is in the expression of 'for you' – in Tatar a pronoun in the dative case, and in Russian a pronoun in the genitive, as required by the preposition *dliia* 'for'. Note that morphologically, the Russian adjectives all express gender (masculine for *rad* 'happy', and feminine for *zdorova* 'healthy' and *schastliiva* 'happy') while Tatar has no grammatical gender.

In examples [3.8] and [3.9] below, I give two examples of fully Tatar equivalents. These utterances were both produced by the speaker in example [3.6] who originally used the code-mixed *chto* 'that'. They were produced by him months after [3.6], when I asked for translation help, claiming that I did not know how say sentence [3.7] in Tatar. The sentences shown in [3.8] and [3.9] were the first two offered by A., who upon a few moment's reflection then offered several more. In fact, all of the speakers cited here who produced code-mixed Russian discourse-pragmatic words proved themselves to be capable of producing a fully Tatar equivalent – if I did not hear them produce the Tatar equivalent either in person or in a recording, I would test them under the guise of requiring translation help. Elicitation by translation often produces calqued constructions, and is advised against by many fieldworkers (e.g., Gil 2001), but in this case, this underhanded sort of elicitation did *not* produce calques, while in fact many of the original code-mixed utterances were calques from Russian. In other words, I did not uncover any competence gaps: speakers who

produced a code-mixed form were also able to produce the standard, textbook Tatar equivalent, or even several equivalent versions.

[3.8] Fully Tatar equivalent of a code-mixed *çbtö* produced by A.

Sin taza-sau häm kör küngel-le bul-gan-ga min (bik) shat.
 you healthy and cheerful mood-with be-PAST.PART-DAT I (very) glad
 ‘I’m happy for you that you’re healthy and in a cheerful mood.’

[3.9] Second fully Tatar equivalent of a code-mixed *çbtö* produced by A.

Sin taza-sau häm kör küngel-le bulgan öchen min(bik) shat.
 you healthy and cheerful mood-with be-PAST.PART for I (very) glad
 ‘I’m happy for you that you’re healthy and in a cheerful mood.’

In both examples [3.8] and [3.9], the (underlined) subordinate clause is an embedded pre-head clause, as opposed to the Russian-style subordinate clause, which comes after the main clause, found in [3.6]. In example [3.8], subordination is realized through a combination of the past participle of the verb *bularga* ‘to be’ and the dative case. This is a similar use of the dative case as seen in [3.6] in the word *singa*, literally ‘to you,’ but meaning something like ‘for you.’ A more faithful, although not particularly English, translation of the sentence found in [3.8] would be ‘I am glad for your being healthy and cheerful.’ In example [3.9], the postposition *öchen* ‘for’ is used in lieu of the dative suffix, with identical semantics. These two means of subordination are more commonly cited as expressing causality and as appropriate translations for Russian *potomu çbtö* ‘because.’ And in this context, the semantics of “I’m glad because you’re happy and cheerful,” and “I’m glad that you’re happy and cheerful” do not seem very far removed. In addition to the two constructions given above, there are eleven other constructions that can be found, either in speech or grammars (Zakiev 1995, 1997;

Kefeli 1999), to express subordination equivalent to Russian *çbtö*. These eleven methods of subordination are listed in Appendix B in the entry for *çbtö*.

Utterances using a code-mixed *çbtöby* ‘in order to, that’ are also realized with a different clausal structure than if produced in Tatar only. As with *çbtö*, a Tatar sentence using the word *çbtöby* as a subordinating conjunction will have a subordinate clause that follows the main clause, Russian-style. Additionally, when *çbtöby* is used, the verb of the subordinate clause is finite. The Tatar equivalents all involve embedded pre-head subordinate clauses, with four different means of subordination, three of which use non-finite verb forms: (1) a verbal noun in *-u* followed by the postposition *öçben* ‘for’; (2) a verb in the *-rga* infinitive followed by the gerund *dip*, literally ‘saying’; (3) a verbal noun in *-u*, possessed and in the dative (so *-uina*), and (4) a verb in the 3rd person singular imperative form, *-sin*.¹⁵

Subordinate clauses using Russian *potomu çbtö* ‘because’ to express causality are syntactically identical to those using the subordinating conjunction *çhönki* ‘because.’ While constructions using *çhönki* are considered to be “pure” Tatar, *çhönki* itself is an earlier function-word borrowing from Persian, a word whose borrowing altered the syntactic realization in utterances where it was used to introduce a subordinate clause. In addition to the construction using *çhönki*, there are 18 analytic constructions in Tatar available for expressing causality, and 10 synthetic constructions. These 28 constructions are also briefly described in Appendix B in the entry for *potomu çbtö*.

¹⁵ All of the suffixes and enclitics described in this chapter have underspecified vowels that are subject to vowel harmony. The 3rd person imperative suffix *-sin* would therefore be more accurately represented as *-sin*. However, vowel harmony and consonant assimilation are not pertinent to the argument of this chapter, so to minimize distraction I am citing suffixes and enclitics in their “unmarked” citation form, with the backed variant of a harmonizing vowel, and without nasal assimilation or devoicing, if applicable.

3.4.1b Discourse structure marking – coordinative

Young Tatar speakers insert the following discourse coordinators when speaking in Tatar-preferred style:

[3.10]

- *tak chto* ‘so’ (coordinating conjunction)
- *i* ‘and’ (coordinating conjunction)
- *tozhe* ‘as well’ (particle)

All three of these coordinators are used as markers of narrative structure, and coordinate “idea units” (Schiffrin 1987) rather than acting as logical operators coordinating items in a list. *And*, the English discourse marker equivalent of Russian *i*, “coordinates idea units and ...continues a speaker’s action” (ibid. 125), which is just how Russian *i* is used when code-mixed in Tatar. Utterances with *tak chto* ‘so’ and *i* ‘and’ are syntactically identical to utterances with their Tatar counterparts, *shunga kürä* and *häm*, respectively. Note, however, that *häm*, like *chönki* discussed above, is an earlier borrowing from Persian; therefore a more Indo-European syntax was already in place. In [3.11] below, we see an example of *tozhe* ‘as well’ produced in Tatar-preferred style.

[3.11] Code-mixed *tozhe*

R.-nı̄ da chakır klub-ka. Anga **tozhe** ber kız tab-ar-sing.
R-ACC also invite-IMP club-DAT him-DAT **also** a girl find-AOR-2PS
‘Invite R. as well to the club. You will find a girl for him **as well**.’ –A.

This utterance is particularly interesting because the speaker has produced the Tatar enclitic *da* in the first sentence to say “also,” but in the second sentence has produced *tozhe* to express the same meaning. In this example, *da* and *tozhe* are used identically in the

constructions. However, in other constructions, while *tozhe* would be used only one time to coordinate two nouns or propositions, *da* will come after both of the nouns or verb phrases being conjoined. This use of *da* can be seen in example [3.21] in section 3.4.1d below.

3.4.1c Discourse structure marking – contrastive

The four contrastive discourse markers code-mixed into Tatar are:

[3.12]

- *ili* ‘or’ (contrastive conjunction)
- *no* ‘but, however’ (contrastive conjunction)
- *a to* ‘else’ (contrastive conjunction)
- *zato* ‘but on the other hand’ (contrastive conjunction)

The contrastive conjunction *ili* ‘or’ is the syntactic equivalent of the earlier Persian borrowings *ya* and *yaki* (both meaning ‘or’); the contrastive conjunction *no* ‘but, however’ is the syntactic equivalent of the earlier Arabic borrowings *läkin*, *ämma*, *bälki* ‘but, however’. Like the English discourse marker *but*, the functional units being linked by these four Russian discourse markers have a “contrastive relationship in either their ideational or interactional content” (Schiffrin 1987: 153). *Ili* ‘or’, like the coordinative conjunction *i* ‘and’, is also used to coordinate functional units and not merely individual items – like its English equivalent *or*, it is used as an option marker in discourse (ibid. 177). This can be seen in example [3.13].

[3.13] Code-mixed *ili*

Sin	Dulkin-ga	shaltirat-öp	kara	äle.	Ili	Kazan-da	ul	tagin
You	Dulkin-DAT	call-2 nd GER.	see	PART.	Or	Kazan-LOC	he	again
kaychan	bulachag-ï-n		belep		kara.			
when	be-FUT.DEF-POSS-ACC		know-2 nd GER.		see-IMP.			

‘Try and call Dulkin [Radio]. **Or** try to learn when he will be in Kazan again.’ -B.

Example [3.14] shows a pragmatically typical use of *no* ‘but, however,’ highlighting the negation of a proposition that might otherwise be expected.

[3.14] Code-mixed *no*

Bu atna-da üzeshchen-när konsert-ï, **no** min jïrla-miy-m.
 This week-LOC amateur-PL concert-POSS **but** I sing-NEG.PRES-1PS.

Kstati anda gel bu sayt-nï – Y.-nï – reklamal-iy-lar.
By the way there always this site-ACC – Y.-ACC – advertise-PRES-3PP

‘This week there is an amateur concert, **but** I am not singing. **By the way**, they are always advertising this site, Y., there.’ –D.

If this utterance had been produced in Tatar only, with *läkin* ‘but’ in lieu of *no* ‘but,’ and *süz unğayında* ‘by the way’ in lieu of *kstati* ‘by the way,’ the syntactic realization would have been identical. Although *kstati* ‘by the way,’ a discourse organizer used for metacommentary on text, is not relevant to the discussion here, I have kept it in the example to demonstrate just how frequent the production of these code-mixed words can be. While in recorded speech produced in Tatar on-stage style, code-mixed Russian words are infrequent, in speech that was merely observed rather than recorded, or in electronic text, like that seen in example [3.14], they can be quite frequent. In an utterance or paragraph of three sentences produced in Tatar-preferred style, it can be possible to find a code-mixed word in each sentence.

3.4.1d Metacommentary and deixis

A variety of Russian words are code-mixed to produce metacommentary on text, some of which is deictic reference to preceding or upcoming information. They are as follows:

[3.15]

- *i vsyo* ‘and that’s it’ (phrase)
- *koroché* ‘in brief’ (adverb)
- *kstati* ‘by the way’ (adverb)
- *naprimer* ‘for example’ (parenthetical word)
- *prichem* ‘moreover’ (conjunction)
- *konechno* ‘certainly, of course’ (adverb)
- *tem bolee* ‘especially, especially as’ (conjunction)
- *to est* ‘that is, i.e.’ (conjunction)

- *chto li* ‘or something’ (particle)
- *neuzheli* ‘really?’ (particle)

- *izvini(te)* ‘pardon, sorry’ (verb in imperative)
- *slushai(te)* ‘listen’ (verb in imperative)
- *davai(te)* ‘come on!’ (verb in imperative)

- *vo!* ‘there’
- *vo!, tak* ‘so, there’
- *nu* ‘so; well’

Because most of these words and phrases are usually found at clausal boundaries, most often at the beginning of a sentence or clause, their use is not associated with a different syntactic realization than found in utterances with their Tatar equivalents. The words code-mixed for metacommentary have the same discourse uses in Tatar as in Russian. Russian *i vsyo!* ‘and that’s it!’ is used to comment upon the preceding proposition and emphasize its force, while *vo!, tak* is used to end topical units (Grenoble 1998: 184). The parenthetical particle *chto li*, loosely translatable as ‘or something’, is most often code-mixed in Tatar as a tag question, and is used, as in Russian, to express uncertainty, doubt, or dissatisfaction on the part of the

speaker, or to highlight the speaker's awareness of the imprecision of the proposition (Merkulova 1998: 336-7). The particle *neuzheli* 'really?' is code-mixed sentence initially, its traditional place in Russian, and just as in Russian it is used to express the speaker's surprise at a just-stated proposition that contradicts something the speaker has held to be true (Baranov and Payar 1998: 301). Like its English equivalent, *izvini(te)* 'pardon, sorry', a verb used in either the informal/singular or formal/plural imperative, is used for distancing from the comment that is about to follow. *Kstati* 'by the way,' (as seen in example [3.14] above) and *slushai(te)* 'listen!' both mark topic shifts, and *kstati*, which is used as a parenthetical introduction to a statement, presents that statement as both autonomus and somehow pertinent to the previously established topic of the discourse (Moro 1998: 249). *Konechno* 'of course,' has two uses in Russian – as a parenthetical adverb used for metacommentary, and as an affirmative particle used to answer a question (and thus an evaluative). Both as a parenthetical adverb and as an affirmative particle, *konechno* is characterized by a simultaneous reference to both the speaker's opinion and external factors that serve as independent guarantors (Kisleva 1998: 345). *Koroche* 'in brief,' like its English equivalent, points out that the speaker is summarizing, and marks equative, or positive structure, such that the content preceding it must approximately conform to the content following it (Shloush 1998: 63) – an example of code-mixed *koroche* 'in brief' can be seen in [3.16] below.

[3.16] Code-mixed *koroche*

Koroche	shaltirat-kach söylesh-äbez.	Poka.
Briefly,	call-PAST.PART speak- PRES.1PP.	Bye.
'In short,	after you call, we'll talk. Bye.' –B.	

Once again, we see a high frequency of code-mixed words. In fact, this example is extracted from the same utterance shown in example [3.13] above – in this utterance of seven sentences, there are four code-mixed discourse-pragmatic words (and no other code-mixed words). While the frequency of code-mixing itself is not unusual, what makes this somewhat remarkable is that this speaker is one of the most culturally and politically active young Tatars to be found in Kazan, a man who has expressed the view that Russians are *bezñeng doşman* ‘our enemy,’ and who has explicitly stated to me on more than one occasion that he avoids speaking Russian unless it is absolutely necessary. He is known among his peers as a man who speaks Tatar beautifully, and was offered to me by several of them as an example of the best Tatar speaker that they know. His home language is Tatar, his college education was in Tatar, and he has worked on Tatar radio and in Tatar print journalism. His command of Tatar on-stage style is fully developed and generally regarded as eloquent – there is none of the faltering, hesitancy, or ungrammaticality that can be seen when other young Tatars speak with a high level of verbal hygiene in a formal register. And even so, this speaker, in a brief utterance, has code-mixed Russian words four times.

In example [3.17] we see a standard use of *to est* ‘i.e.’, that is, preceding a restatement or clarifying remark:

[3.17] Code-mixed *to est* ‘that is’

P.,	bez	bit	Mishär	kız-lar-ı	tügel.	Mishär	tel-e-n,
P.	we	EMPH.PART	Mishar	girl-PL-POSS	not.	Mishar	language-POSS-ACC
to est		dialekt-ı-n		angla-miy-biz			bit.
that is		dialect-POSS-ACC		understand-NEG.PRES-1PP			EMPH.PART.

‘P., we aren’t Mishar girls. We really don’t understand the Mishar language, **that is**, its dialect.’ –E.

Quite similar to *to est*’ is the use of *naprimer* ‘for example,’ which just like its English equivalent signals that the words to follow are meant to illustrate the topic at hand, as seen in example [3.18]:

[3.18] Code-mixed *naprimer* ‘for example’

Ul	chinlapta	burzhuy.	Min	naprimer	Idel	tämäke-se-n
he	in reality	bourgeois	I	for example	Idel	tobacco-POSS-ACC
tart-a-m,		ä	ul	Parliament	tart-a.	
smoke-PRES-1PS	and/but	he	Parliament	Parliament	smoke-PRES	

‘He really is a bourgeois! I, **for example**, smoke Idel tobacco, and he smokes Parliament.’
–B.

The Russian discourse particle *vot* has been analyzed as a modal particle without semantics, “best understood in terms of [its] discourse and pragmatic functions” (Grenoble 1998: 21). It is a “presentative” deictic equivalent to a verbal gesture, and directs attention either forward or backward in a text. This can be seen in example [3.19], in which *vot* is backward-looking.

[3.20] Code-mixed *vot*

Vot	singa	tipychnyi	kön.
There	you-DAT	typical	day

“So **there** you have a **typical** day.” –G.

This statement concludes a paragraph in an e-mail where G. has been describing a typical day at his workplace, essentially listing activities hour by hour. This usage of deictic *vot* is exactly parallel to its Russian use, and also parallel to the English use of ‘there’, or ‘so there’ to point to information in the discourse. Note that the use of Russian *tipychnyi* for ‘typical’ in lieu of Tatar *tipik* or *gadättäge* is not representative of the speaker’s usage – he is echoing the

language of my question, in which I had forgotten the Tatar word for ‘typical’ and, as most less-competent Tatar speakers would, had substituted the Russian adjective. This speaker is quite careful about his Tatar speech, even though in a professional context he uses Russian only: this is the only example I have of a Russian adjective, complete with Russian derivational morphology, for all of his speech and writing.

The Russian particle *nu* is used in both Russian and Tatar to “introduce a new topic or signal the continuation of a previously established, activated topic” and can also be used to express a conclusion, summary, or influence (Grenoble 1998: 181, 183). *Nu* is quite frequent in both Russian and Tatar, and can be found opening several conversational turns in a row – it is loosely equivalent to the English discourse particles *so* and *well*, as can be seen in example [3.20]. In this conversation, speaker H. has been telling speaker I. about a journalist of mixed ethnicity, half-Tatar and half-Russian, whose Tatar language skills are not the best.

[3.20] Code-mixed *nu* ‘so’

H. ... menä shundiŷ üze tatar-cha söyläsh-i. Ä üze
 ...like this himself Tatar-ADV speak-PRES.3PS And himself

menä tatar zhurnalistika-sin-da.
 there Tatar journalism-POSS-LOC

I. **Nu,** ul yaz-a-mi song üze?
 So, he write-PRES.3PS-?PT after all himself

Tatar-cha yaz-a-mi?
 Tatar-ADV write- PRES.3PS-?PT

H. ...he speaks Tatar like this. And there he is in Tatar journalism.

I. **So,** he himself writes after all? Writes in Tatar?

Here *nu* is marking both the continuation of the topic and the introduction of a summary of that topic.

The discourse-packaging Russian imperative *slushai* ‘listen!’ is a directive that both marks and draws attention to a topic change – it is used to select both the next speaker and the next topic (Grenoble 1998: 157). Example [3.21] shows, among other things, a code-mixed *slushai* ‘listen!’ as directing attention to a topic change and marking F. as the next speaker in the conversation. The immediately preceding sentences have been about whether or not the conversation participants will be attending the club meeting on that coming Saturday.

[3.21] Code-mixed *slushai*

A: **Slushai**, sin gitara-ng-mï küter-ep kil-ä al-miy-sing-mi?
listen you guitar-2PS.POSS-?PT carry-2ndGER come-1stGER take-NEG.PRES-2PS-?PT

F: Gitara...
 gitar

A: **Slushai**, bik küp zhurnalist-lar kil-ä. ORT, TRT belän
 listen very many journalist-PL come-3PS.PRES. ORT TRT with
 kil-ergä dä **mozhet**, kil-mäskä dä **mozhet**.
 come-INF also **might** come-NEG-INF also **might**.

A: **Listen**, can you bring your guitar?

F: Gitar...

A: **Listen**, a lot of journalists are coming. ORT and TRT [television stations] **might** come with them, and they also **might** not come.

The Tatar equivalent of Russian *slushai* is *kara äle*, the informal/single imperative form of the verb ‘to see’ followed by an emphatic particle. (This particle *äle* is often translated into Russian as the particle *-ka*, so that the literal Russian equivalent of *kara äle* would be *smotri-ka*, ‘just look!’). Note that this speaker, A., also uses this Tatar equivalent, as seen in example

[3.42] below. Verbs of both looking and listening are common discourse markers of topic shift, and indeed Tatar *tingla äle* ‘listen!’ could be used in this context, although it does seem to be used with less frequency.

There are three other items of interest in this example. The first is the fully Tatar double use of the *da* enclitic (here realized as *dä* due to rules of vowel harmony) to coordinate two verb phrases: the television crews might come with the journalists, and they might not come. This is the traditional, prescriptively recommended use of *da*, produced by the same speaker who gave us code-mixed *tozbe* ‘also’ in example [3.8] above. The second item of interest is the code-mixing of Russian *mozhet* as an epistemic (epistemics in general, and *mozhet* in particular, are described in section 3.4.3d below). Finally, we once again see a high frequency of code-mixed words: in two conversational turns, a total of three sentences, the speaker code-mixes four times.

3.4.2 Interactional performatives

The second major category of code-mixed discourse-pragmatic words is that of interactional performatives. There are four such Russian words used in Tatar-preferred style:

[3.22]

- *privyet* ‘hi’
- *poka* ‘bye’
- *pozhaluista* ‘please’
- *izvinit(te)* ‘sorry, excuse me’

As might be expected, there are Tatar equivalents for all of these words, equivalents that are used constantly. The Tatar equivalent of Russian *privyet* ‘hi’ is *isänme* (singular) and *isänmesez* (plural), or, alternatively, *sälam*, itself a borrowing from Arabic. Interestingly, I have yet to

find an instance of code-mixed Russian *zdravstvui(te)*, the more formal Russian greeting. This is most likely because Tatar-preferred style is generally in informal register, and all of the utterances from which my data were taken were produced in circumstances in which the use of *zdravstvui(te)* would be potentially inappropriate. The same holds true for leave-taking: only the casual *poka* is code-mixed in lieu of Tatar *sau bul* (singular) or *sau buligiz* (plural), and we do not find the more formal *do svidaniia* ‘goodbye’. Young Tatars will often double-mark their leave-taking, first saying goodbye in Tatar and then in Russian: “*Sau-bul! Poka!*” This seems to suggest that the Russian word is somehow necessary to the leave-taking, that the act of leave-taking has not been performed until it has been performed in Russian.¹⁶

Perhaps the same is true for the use of Russian *pozhaluista* ‘please’ rather than Tatar *zinhar* or *zinhar öchen* (or, in the case of imperatives, a *-chi* politeness suffix on the end of the verb), where politeness is emphasized by means of the force associated with Russian. However, neither of the Tatar expressions of politeness are really equivalent to Russian *pozhaluista*: *zinhar* is marked for formality, and is found in the speech of Tatars striving for the literary standard, while an imperative with the *-chi* politeness suffix is more forceful than a request with Russian *pozhaluista*. *Izvini(te)* ‘pardon, sorry,’ here a performative resulting in an apology of some kind (as opposed to a distancing discourse marker, as seen above in section 3.4.1d) may also have additional pragmatic force due to its expression in Russian, rather than Tatar *gafu it* (singular) and *gafu itegez* (plural). For all of the words in this category, “their

¹⁶ Similar behavior can be found among adult Native American second-language learners in revitalization immersion programs, speakers who have made a commitment to speak only this Native American language with their language mentors, but who find themselves unable to hang up the phone if the conversation has ended only with minority-language words of leave-taking – the conversation is not felt to have ended until the word “Goodbye” has been uttered on both sides (Leanne Hinton, personal communication).

conventional meaning *is* their pragmatic function. These types of words are almost without semantics” (Eve Sweetser, personal communication).

3.4.3 Evaluatives

The third and largest pragmatic category of code-mixed words is that of evaluatives. In this category, we find words of positive and negative evaluation, hedges, epistemics (i.e., evaluations of probability), and evaluations of time and degree.

Also code-mixed is the one evaluative Russian question word, *kak* ‘how’. None of the other Russian question words (e.g., *kto* ‘who’, *kogda* ‘when’, *kakoi* ‘which’) are code-mixed in Tatar-preferred style. Additionally, *kak* is only used in its (grammaticalized) evaluative meaning, never in the root meaning that refers to manner. That is, a speaker might use a code-mixed *kak* in a question like, “How did you like Liliya?” but would not use *kak* in a question like, “How do you make those blini again?” Usually the evaluator is explicit and in the dative case, just as in Russian, although this is not necessarily a sign of calquing, as a dative experiencer in constructions of this sort is quite common. If the evaluator is not explicit, it is an implied “you,” as seen in example [3.23].

[3.23] Code-mixed *kak*

Khäter Kōn-e **kak?**
Memory day-POSS **how?**
‘How was Memory Day?’ –J.

As described in Chapter One, *Khäter Kōne* ‘Memory Day’ is a holiday that has been observed in Tatarstan each October since the fall of the Soviet Union. It commemorates the fall of the Kazan Khanate to Ivan the Terrible in 1552, when Tatars fell under Russian political domination. The day is an emotional one for many, and the observance is highly politicized

and dominated by Tatar nationalists, some of them extremists. In fact, a number of Tatars were jailed in October of 2001 for speeches they had made during the Memory Day gathering, speeches that were labeled “seditious” by the government. It is particularly telling, and a mark of the lack of awareness of this code-mixing behavior, that this club member has asked for an evaluation of arguably the most anti-Russian of events to take place annually by using a Russian word.

3.4.3a Positive evaluation

All of the discourse-pragmatic words code-mixed to express positive evaluation are interjections, and thus disjunct constituents unrelated to the syntax of the surrounding discourse. They are:

[3.24]

- *chudo* ‘cool!’ (lit. ‘miracle’)
- *molodets* ‘well done!’
- *prikol’* ‘how funny!’
- *tochno* ‘precisely, exactly’
- *voobshche klassno* ‘totally cool’

Example [3.25] shows a typical use of *tochno* ‘exactly’ when used for positive evaluation.

[3.25] Code-mixed *tochno*

SW: Yuk, beraz dorfa, minemchä.
no, little rude, in my opinion.

A: **Tochno.** ‘Käbestä ech-e-n-dä’ bul-sa, nechkä-räk bul-ir
exactly. cabbage inside-POSS-LOC be-COND, subtle-COMP be-AOR

i-de, yomshag-rak bul-ir i-de.
be-PAST.DEF refined-COMP be-AOR be-PAST.DEF.

SW: No, a little rude, I think.

A: **Exactly.** If it had been ‘inside a cabbage,’ it would have been more subtle, it would have been more refined.

Several of these interjections are sociolinguistically marked as youth speech: they are *chudo* ‘cool’, *prikol* ‘how funny!’, and *voobsbche klassno* ‘totally cool.’ In this respect, they can be seen in some ways as filling a gap in Tatar, which does not appear to have any slang. In my months of observation and interviews, I observed and was informed of only one word of slang, Tatar *tikä*, literally ‘steep,’ used to mean ‘cool.’ However, not only is this word used only by members of very small social network (a small subset of club members), it is actually a self-conscious, and in some ways self-mocking word, because it is a calque of a Russian slang word meaning ‘cool’, *krutoi/krutaia/krutoie* (m/f/n), which literally means ‘steep.’ It is not a slang word that has been generated in the usual way. One club member, when asked if he knew any Tatar slang that he could teach me, said, “No, there’s no slang in Tatar. Maybe in the villages, but not here. You have to use a language all the time in order to come up with slang.” Implied in this comment is that when young Tatars need to use slang or expressions marked as youth speech, they simply turn to Russian (or perhaps are already speaking in Russian if it is a situation where slang use would be appropriate).

3.4.3b Negative evaluation

The two code-mixed words of negative evaluation are also interjections:

[3.26]

- *uzhas* ‘how awful!’
- *zhalke* ‘it’s a shame’

Neither of these words are sociolinguistically marked, and they can be used by speakers of all ages.

3.4.3c Hedges

Six Russian adverbs are used as hedges in Tatar:

[3.27]

- *chut* ‘almost’
- *kebot* ‘only’
- *okolo* ‘around’
- *pochti* ‘almost’
- *prosto* ‘simply, just’
- *sluchaino* ‘accidentally, by chance’

The adverb *prosto* ‘simply, just’ is used as a “minimizing” hedge in Russian, one that shows that the proposition does not contain “any kind of additional augmentation imposed on it by the situation or context” (Baranov et al. 1993: 171), as can be seen in example [3.28]:

[3.28] Code-mixed *prosto*

Minem	fiker-em-che,		bezgë	(latin-ni	yaklau-chi-lar-ga)	prosto	
My	thought-1PS.POSS-ADV,		we-DAT	(latin-ACC	defending-AGT-PL- DAT)	just	
sabir	bul-irga	kirëk,	äkren	genä	latin-ni	gamäl-gä	kert-ergä,
patient	be-INF	necessary,	slow	only	latin-ACC	usage-DAT	enter-INF,
a	kart	babay-lar	10-15	yıl-da	ülep	bet-ergä	tiyesh-lär.
and	old	grandpa-PL	10-15	year-LOC	die-2 nd GER	end-INF	should-PL.

‘In my opinion, we (the defenders of the Latin alphabet) **just** need to be patient, Latin will only enter into use slowly, and the old grandpas should all be dead in 10-15 years.’ –E.

Had the Tatar equivalent, *tik* ‘just, simply’, been used, the morphosyntactic realization would have been identical.

Like the interrogative *kak* ‘how’, *okolo* ‘around, approximately’ is code-mixed only in its grammaticalized and evaluative meaning: there are no examples of code-mixed *okolo* with its root spatial meaning of (literally) ‘around.’ Additionally, the code-mixing of the hedges *okolo* ‘around, approximately’ and *pochti* ‘almost’ comes along with a composite syntax: these

adverbs, both when used in Russian and when code-mixed in Tatar, precede the word or phrase that they are modifying, while their Tatar equivalents follow the modified word or phrase. The hedge *pochti* ‘almost’ introduces a proposition that “the speaker places in the context R (= ‘it is not possible to say that it is R, but it is close to R’), and suggests that the several ways in which it differs from R can be disregarded” (Baranov et al. 1993: 62). This can be seen in example [3.29].

[3.29] Code-mixed *pochti*

Shifali	su	(mineral	su)	faydali-rak.	Anda	mineral	toz-lar
curative	water	(mineral	water)	useful-COMP	There	mineral	salt-PL
häm	pochti	nol	kaloriya	genä.			
and	almost	zero	calorie	PART			

‘Curative water (mineral water) is more useful. It has mineral salts and **almost** no calories.’ –A.

The equivalent Tatar form *diyärlek*, would come after the modified phrase, at the sentence’s end.

A more dramatic example of composite syntax can be seen in the sentence shown in [3.30], involving a code-mixed *chut* ‘almost.’ While in Russian *chut* can be used to qualify a variety of words, including verbs, adjectives, and adverbs, and can mean either ‘hardly’ or ‘almost,’ when code-mixed in Tatar, *chut* is only used to modify verbs in a construction meaning that the action of the verb almost took place, but in the end did not take place.

[3.30] Code-mixed *chut'*

Bashta **chut'** yokla-p kit-mä-de-k, yariy äle
At first **almost** sleep-2ndGER leave-NEG-PAST.DEF-1PP ok EMPH PART

zaman-nar dürt jir jirla-p uya-t-ti-lar.
time- PL four song sing-2ndGER wake-CAUS-PAST.DEF-3PP.

'At first, we almost fell asleep, ok, after singing four songs they woke (us) up.' –E.

When used in Russian in this sense, *chut'* requires a negative verb:

[3.31] Russian equivalent

My chut' ne zasnu-l-i.
We almost NEG fall.asleep.PERF.PAST-PL
'We almost fell asleep.'

The speaker of [3.30] has used a negative verb form as required by *chut'*: the finite auxiliary verb *kitmädek* is in the negative (the positive form of the verb, in appropriate tense and person, would be *kittek*). The remainder of the construction's syntax is governed by the requirements of *kitergä* 'to leave' when used as an auxiliary verb. When used with a converb in the 2nd gerund form (ending in *-ip*, *-ep*, or *-p*), the grammaticalized *kitergä* gives one of the following meanings, all of which are clearly derived from its root meaning of 'to leave': (1) action away from a point or the speaker, (2) acquisition of some sort of quality, (3) the beginning of an action, or (4) impulsive or sudden action. (Zakiyev 1997: 188). Here, *kitergä* is being used in its third sense, to mark the beginning of an action, and the periphrastic verbal form *yoklap kit-* means 'fall asleep.'

Now let's look at what the fully Tatar equivalent of [3.30], shown in [3.32], where the construction equivalent to 'we almost fell asleep' is underlined:

[3.32] Tatar equivalent

Bashta	yokī-ga	kit-ä	yaz-dī-k,	yariy	äle
at first	sleep-DAT	leave-1 st GER	miss-PAST.DEF-1PP	ok	EMPH.PART
zaman-nar	dürt	jir	jir-la-p	uya-t-ti-lar.	
time-PL	four	song	sing-2 nd GER	wake-CAUS-PAST.DEF-3PP	

‘At first we almost fell asleep, ok, after singing four songs they woke (us) up.’

Here we can see several differences from the construction in [3.30]. The first is that there is no separate word meaning ‘almost’. Instead, the semantics of the action almost happening are expressed by the periphrastic combination of the archaic verb *yazırğa* ‘to miss’ (now only used in this auxiliary form, and commonly reinterpreted as present-day *yazırğa* ‘to write’) and a converb in the first gerund form (ending in *-a*, *-ä*, or *-iy*). Note that the verb is *not* in the negative, as there is no requirement for a negative. Additionally, the collocation usually used for ‘to fall asleep’ is not a periphrastic verb with an auxiliary of *kitergä*, but rather the noun *yokī* ‘sleep’ in the dative case in combination with the verb *kitergä*, literally, ‘to leave into sleep.’

Therefore, the utterance produce in [3.30] can be seen as a composite of the two languages on several levels: lexical (with the use of the Russian word *chut*), morphological (due to the negative morpheme required by the use of *chut*), and syntactic, as the syntax is neither fully Russian (and not a calque), nor fully Tatar.

3.4.3d Epistemics

Young Tatars use two Russian words (one verbs and one adverb) as epistemics, that is, to evaluate probability:

[3.33]

- *mozhet* ‘might, maybe, possibly’ (verb)
- *navernoe* ‘probably, most likely’ (adverb)

The verb *mozhet* appears in the third-person singular form only. An example of code-mixed *mozhet* was seen in [3.21] above, and here is another:

[3.34] Code-mixed *mozhet*

Ike-nche-dän, min **mozhet** yagimli tügel-der. Kaydan bel-ä-seng?
Two-ORD-ABL I **might** sweet not-INDEF.PART From where know- PRES-2PS?
‘In the second place, I **might** not be sweet. How would you know?’ –D.

The Tatar equivalents of *mozhet*, *bälkei* and *mögaen*, are adverbs. They are usually at clausal boundaries, most often at a clause beginning. Code-mixed *mozhet*, on the other hand, seems to stay adjacent to the word it is modifying, with at most an intervening enclitic (as seen in [3.21]). The verb *moch* ‘to be able’ is only code-mixed in this epistemic usage, and in the fixed third-person singular form: it is never used in its root meaning as a conjugated auxiliary verb expressing ability.

3.4.3e Evaluations of time and degree

In general, code-mixed words that are used as evaluations of time and degree are not associated with a composite or alternate morphosyntactic realization. This is, perhaps, because both the Russian words and their Tatar equivalents (which are often two-word phrases where the second word is an enclitic) tend to precede the words or propositions

being modified, and it is when Tatar equivalents are not pre-head that we usually find a composite syntax (i.e., partly Tatar and partly Russian word order) associated with the code-mixing.

The code-mixed evaluations of time are all adverbs:

- *uzbe* ‘already’
- *skoro* ‘soon’
- *srazu* ‘immediately’
- *srochno* ‘quickly, urgently, soon’

Example [3.35] shows a typical use of *skoro* ‘soon.’

[3.35] Code-mixed *skoro*

Böten	khalik-tan	ser	it-ep	sakl-iy-lar.	Läkin	minga
entire	people-ABL	secret	do-2 nd GER	preserve- PRES-3PP	but	me-DAT
informer-lar	jitker-de-lär.	Skoro	min	dä	Moskva-ga	bar-miy-m.
informant-PL	lead-PAST.DEF-3PP	soon	I	also	Moscow-DAT	go-NEG.PRES-1PS

‘They keep secrets from the entire nation. But they led informants to me. **Soon** I too will not go to Moscow.’ –B.

The Tatar equivalent, *tizdän*, also an adverb, would be used identically.

Example [3.36] shows a typical use of *srochno* ‘quickly, urgently’:

[3.36] Code-mixed *srochno*

I. pochta yashchig-ïng-ï ach-ïp kara äle. Anda min singa
 I. mail box-2PS.POSS-ACC open-2nd GER look-IMP PART there I you-DAT

khat yaz-gan i-de-m. Mömkin bul-sa **srochno** gïna
 letter write-PAST.PART be-PAST-1PS. Possible be-COND **urgently** PART

jawab bir-che.
 answer give-IMP-POL

‘I, look at your mailbox. I wrote you a letter there. If it’s possible, please answer **quickly.**’ -D.

Note that the *srochno* has been fully integrated into the Tatar syntax and is modified by the Tatar emphatic particle *gïna* rather than by the Russian emphatic particle *эше*, which, interestingly enough, one does not find code-mixed in Tatar-preferred style. The Tatar equivalent, *achigip*, although more literal in meaning than Russian *srochno* and thus not entirely equivalent, would be found in a sentence with the same morphosyntactic realization. While *achigip* is a gerund, as opposed to an adverb (*srochno*’s grammatical class), the line between Tatar gerunds and adverbs is faint and often artificially drawn.

The pragmatic category that contains the most code-mixed Russian words is that of evaluations of degree. These words represent more grammatical classes than can be found in the other categories listed above, but are united by their pragmatic function. The first subset of these evaluators of degree are words that evaluate the level of the characteristic quality being expressed:

[3.37]

- *daxbe* ‘even’ (particle)
- *eshche* ‘more, still’ (adverb)
- *slishkom* ‘too, overly’ (adverb)

While utterances with one of the above words have the same syntactic realization as if the Tatar counterpart had been used, the particle *dazhe* ‘even’ is sometimes used together with the Tatar particle *kbätta* ‘even’ to double-mark the modified word in the construction ‘*dazhe kbätta X*’, meaning ‘even X.’ However, the *dazhe* seen in example [3.38] is used alone:

[3.38] Code-mixed *dazhe*

Bel-ä-seng-me, know-PRES-2PS-?PT	min I	moning now-GEN	khätle küp up to much	it-ep, do-2 nd GER
dazhe khat-ta even letter-LOC	yaz-gan-im write-PAST.PART-1PS.POSS		yuk. not	

“You know, up to now I’ve been really busy, I haven’t **even** written a letter.” –K.

The second major grouping of these words of degree are all used to somehow evaluate the relationship to reality or the general applicability of the proposition in question. At the most extreme end are the words *da* ‘yes’ and *konechno* ‘of course’, affirmative particles used to express agreement. The other words and phrases used to evaluate applicability are:

[3.39]

- *absolutno* ‘absolutely’ (adverb)
- *inogda* ‘sometimes’ (adverb)
- *kak vseгда* ‘as always’ (phrase)
- *poluchaetsia* ‘(it) turns out’ (verb)
- *na samom dele* ‘in reality, in fact, actually’ (phrase)
- *ni v koem sluchae* ‘by no means’ (phrase)
- *obiazatel’no* ‘absolutely, without fail’ (adverb)
- *obychno* ‘usually’ (adverb)
- *sovsem* ‘totally, completely’ (adverb)
- *tochno* ‘exactly’ (adverb)
- *voobshche* ‘in general’ (adverb)
- *v obshchem* ‘in general’ (phrase)

Russian *absolutno* ‘absolutely’ has no Tatar equivalent, and while this adverbial form is not a standard borrowing, and can thus be considered to be code-mixed, the Tatar adjective *absolut* is in fact a standard borrowing taken from Russian *absolutnyi* (most Russian adjectives ending in *-nyi*, a adjectival suffix cited here in masculine form, were borrowed into Tatar without the *-nyi* suffix).

In example [3.40], we see a typical use of Russian *obychno* ‘usually,’ whose placement, like its Tatar equivalent *gadettä*, is reasonably free within a clause, although it is more commonly used as one of the first words of the clause or sentence, thus immediately contextualizing what is about to follow:

[3.40] Code-mixed *obychno*

Kiz-lar	tämle	äyber-lär	peshher-ep	kil-ä-lär.	Ä	min	obychno
Girl-PL	tasty	thing-PL	cook-2 nd GER	come PRES-3PP.	And I		usually
kosh	tel-e		kiter-ä-m...				
bird	tongue-POSS		bring-PRES-1PS				

‘Girls come having cooked tasty things. And I **usually** bring bird’s tongue [a type of pastry]...’ –C.

The Russian adverb *voobshche* ‘generally’, similar in meaning to *obychno* ‘usually’, has three discursive uses: (1) it expresses the relationship of a specific proposition to general conditions, such that it can often be translated as ‘always’; (2) it highlights the difference between an exception and a rule; (3) it is used to generalize from a specific case (Baranov et al. 1993: 105-111). The phrase *v obshchem* ‘in general’ is used by the speaker to introduce a specific proposition or characteristic that is the most useful representative of a more general situation (ibid: 119), as seen in example [3.41]. These words are often used in Tatar, just as in Russian, to show that reported speech is a summary rather than a direct or complete

citation, and the combination *i voobshche* ‘and in general’ usually precedes either a conclusion or a summarizing statement (Grenoble 1998: 121, 182).

[3.41] Code-mixed *v obshchem*

Yarat-kan	jir-chi	eis of beis,	doktor alban,	v obshchem	min
love-PAST.PART	song-AGT	Ace of Base	Doctor Alban	in general	I
jir-ga	kara-p	yarat-a-m.			
song-DAT	look-2 nd GER	love-PRES-1PS			

‘My favorite singers are Ace of Base, Doctor Alban – **in general** I love to check out music.’ –G.

When adverbs evaluating applicability are utterance-final, as seen in example [3.42], emphasis is being placed on the evaluative word.

[3.42] Code-mixed *inogda*

Kara	äle,	sin	uk-iy-sing-mi	intertat-ni	inogda?
look-IMP	EMPH.PART	you	read-PRES-2PS-?PT	intertat-ACC	sometimes

‘Listen, do you **sometimes** read Intertat [a Tatar internet newsite]?’ –A.

Finally, we have the phrase *na samom dele* ‘in fact, actually’, which is quite commonly used in Russian, and points out that the speaker is departing from his or her own representation of reality and is registering the status of a previously stated proposition or situation as in fact real (Baranov et al. 1993: 92).

[3.43] Code-mixed *na samom dele* ‘in reality’

Ä	khätta	na samom dele	sikher	bar.
and/but	even	in reality	magic	is

‘But there is magic even **in reality**.’ –G.

This example actually shows a somewhat atypical use of *na samom dele*, an expression more often used to contrast the difference between a hypothetical situation and a reality that is worse. Even so, its discourse-pragmatic functionality remains the same.

3.5 Russian words that are *not* code-mixed in Tatar-preferred style

If we exclude from consideration nouns and verbs, which are a tiny percentage of the Russian words code-mixed in Tatar-preferred style and clearly belong to a different class than the code-mixed words seen in section 3.4, then the data I have presented above are not merely a selected subset of the code-mixed words found in my Tatar corpus. They are, in fact, *all* of the code-mixed words produced by these selected Tatar speakers. Every single one of them, without exception, fits into one of the three discourse-pragmatic categories listed in section 3.4. This is the first major piece of evidence that the organizing principle of discourse-pragmatic functionality is a valid one.

Using discourse-pragmatic functionality to classify these code-mixed words leads to the prediction that non-content words and morphemes that are without discourse-pragmatic function will not be code-mixed in Tatar-preferred style, and that is indeed the case.

Prototypical “grammatical” non-content words and morphology include:

- Pronouns
- Interrogatives
- Prepositions
- Numbers
- Auxiliary verbs
- Tense and aspect markers
- Verbal and nominal inflectional morphology

Young Tatars do not code-mix any pronouns in Tatar-preferred style: we do not find personal pronouns, possessive pronouns, interrogative pronouns, demonstrative pronouns, or general pronouns (in the sense of ‘all’ or ‘everybody’). Interrogatives too are not code-mixed, with the one exception of *kak* ‘how,’ as described above in section 3.4.3. Recall that *kak* as an interrogative is only used in its evaluative sense, and never to ask about manner. Additionally *kak* is never code-mixed in any of its other grammatical functions, such as a subordinating conjunction meaning ‘as’ or as a conjunction meaning ‘both’ (when used with the adverb *tak* ‘so’). Of the 29 Russian prepositions, only *okolo* is code-mixed, and never as a preposition in its root spatial meaning, but only in its grammaticalized adverbial form meaning ‘approximately,’ used as a hedge. Numbers too are not code-mixed.

As for auxiliary verbs, only *moch* ‘to be able’ is code-mixed, and never in its deontic sense, expressing ability. Instead, all instantiations of code-mixed *moch* are the fixed third-person singular form *mozhet*, used only for epistemic evaluation of possibility. The auxiliary verb *byt* ‘to be’ is not code-mixed as a past tense imperfective auxiliary (*byl* m., *byla* f., *bylo* n., *byli* pl.) nor is it code-mixed as a future tense imperfective auxiliary (*budet*). Finally, only free morphemes are code-mixed, while bound morphemes are not – including tense and aspect markers (defined by Myers-Scotton (1993) and Fuller (1996) as “prototypical” non-content morphemes) and inflectional morphology and derivational morphology. Less-competent speakers when performing in Tatar on-stage style will sometimes use the Russian *-skii* derivational morpheme to create an adjectival phrase. This is most likely due to a lack of competence in the use of *izafet*, the Tatar construction used to link two nouns so that one modifies the other: for example, an Tatar architect without a developed professional register, for whom Tatar is a home language only, used the expression *Bolgarskii stil* ‘Bolgar style’

several times in a presentation at the Tatar Social Club. *Bolgarskii stil* is an adjectival phrase, as the derivational suffix *-skii* has been added to the noun *Bolgar* to derive an adjective. The Tatar equivalent is a compound noun phrase using *izafet*: *Bolgar stile* ‘Bolgar style’, two nouns, with the second noun *stil* having the *izafet* suffix *-e*, also analyzed as a possessive, showing its relationship with the preceding noun. However, this use of derivational morphology, or of bound morphemes at all, is not found among the Tatar-dominant or balanced bilinguals from whom the data in this chapter were taken.

3.6 Conclusions

We have seen that an understanding of local practices and norms – here, the Tatar “discourse of purity” – is vital for the analysis of linguistic performance. The style shifting of the young Tatar bilinguals cited in this chapter – for whom an important part of their Tatar identity is their command and frequent use of “pure” Tatar in a variety of registers and domains – highlights their lack of awareness of their code-mixing of Russian discourse-pragmatic words, words that are not present when they perform at the highest level of verbal hygiene in Tatar on-stage style. It was only by both gaining access to and then delineating the differences between Tatar on-stage style and Tatar-preferred style that I was able to understand the systematic use and significance of code-mixed Russian discourse-pragmatic words discussed here.

An analysis of previously borrowed Russian function words showed that the borrowing of discourse-pragmatic words is not a new phenomenon in Tatar. Discourse-pragmatic functionality was seen to be the key to the categorization of Russian words that are code-mixed in Tatar-preferred style, all of which belong to one of three pragmatic

groupings: (1) markers of discourse structure and force, (2) interactional performatives, and (3) evaluatives. Further proof that discourse-pragmatic functionality is the general organizing principle for the code-mixing found in Tatar-preferred style takes the form of negative evidence, where non-content words and morphemes that do not have pragmatic function are not code-mixed.

In Chapter Four, “The role of discourse-pragmatics in lexical and structural borrowing,” I will lay out a new model of lexical and structural borrowing, one that accounts for both the patterning of discourse-pragmatic borrowings and the structural alterations that their code-mixing or borrowing brings about. The Tatar data analyzed in this chapter, which included phenomena neither described nor accounted for by any of the current discussions or theories of language shift, will be contextualized with cross-linguistic data from other comparable language contact situations. Distinguishing between the denotative and metalinguistic levels of discourse will prove to be essential to both recognizing and understanding one pathway of structural influence from a dominant language and language attrition in a minority language.

Chapter Four

The role of discourse-pragmatics in lexical and structural borrowing

4.1 A new model of lexical and structural borrowing

In Chapter Three I presented a newly discovered pathway of language attrition, the code-mixing of dominant-language words in speech that is otherwise free of code-switching. Discourse-pragmatics was seen to be the key to understanding the motivations for and mechanisms of the code-mixing, and all of the code-mixed words had metalinguistic functionality: they structured discourse, expressed the speaker's evaluative stance, and served as interactional performatives. This code-mixing of discourse-pragmatic words is associated with an innovated composite morphosyntax such that Tatar is becoming more like Russian on both the lexical and morphosyntactic levels. In this chapter, I will show that discourse-pragmatics is central to understanding both lexical and structural borrowing in contact situations similar to that of present-day Tatar and Russian.

Drawing on data from Tatar under the influence of Russian and from a variety of other language contact situations, I present here a new model of lexical and structural borrowing, a model where the demarcation between the denotative and metalinguistic levels of language is central to understanding both the motivations for and patterns of borrowing. The languages presented in this chapter that have followed this pathway of contact-induced change have certain features in common, and this model is not necessarily applicable to all language contact situations. The first is that the language that is being lexically and structurally influenced has either limited or contracting functional domains, and there is at least one high-prestige functional domain that has been ceded to the dominant language and

the dominant language is associated with either power or prestige, or both (see section 3.1 of Chapter Three for a description of contracting functional domains). Second, some of the languages undergoing contact-induced change will have speakers with language ideologies that emphasize linguistic “purity,” which appears to throw unconscious code-mixing into relief. Third, in order to reach later stages of lexical and structural borrowing, there must be a portion of the populace that is highly competent in the dominant or superstrate language. Finally, there is often a contracting population of speakers of the language that is being influenced, i.e., it is a language undergoing multi-generational language shift (called “gradual death” in Campbell and Muntzel’s 1989 typology of language death situations).

It is well known that in language contact situations, certain types of words will be borrowed before others, and certain structural changes will occur before others – these hierarchies of borrowing phenomena have been associated with intensity of contact (e.g., Thomason and Kaufman 1988) or a change in the matrix language in code-switching (Myers-Scotton 1993, 1998). For example, in their well-known borrowing scale, Thomason and Kaufman note that content words are borrowed before function words, and that function words such as conjunctions and adverbial particles are borrowed before function words such as adpositions. In addition, they demonstrate that in “casual contact” situations, there will not be structural borrowing, while in “more intense” contact situations (which are associated with function word borrowing) we find structural borrowing of phonological, syntactic, and lexical semantic features, where some borrowing phenomena precede others. For example, at level 3 of borrowing in the scale, we may find small changes in word order, such as borrowed postpositions in a language that is otherwise prepositional, while “fairly extensive word order changes” will not be found until level 4 of borrowing, which is associated with “strong cultural pressure” (Thomason and Kaufman 1988: 74-75).

While this borrowing scale aptly typologizes the lexical and structural borrowing found in contact situations of varying intensity, it does not provide real motivation for the hierarchy of borrowing it presents. Current models of contact-induced change separate the lexicon into two classes, content words vs. system words and morphemes. However, there are actually three classes of words that play a role in borrowing: content words, discourse-pragmatic words, and system words. In fact, it is the demarcation between denotative language (usually referred to in language contact studies as “content words”) and metalinguistic language (previously unrecognized as a class, and thus included in the class referred to as “function words” or “grammatical words”) that is key to understanding which words are borrowed in what order, and why. Using discourse-pragmatic functionality as an analytical basis gives a model with three stages of lexical borrowing, where each stage necessarily precedes the next:

- **Stage One** does not require a high number of bilingual speakers in the populace, nor great proficiency in the dominant language. In this stage we find content-word borrowings only – mostly nouns and verbs – and no other borrowing phenomena. This stage is essentially equivalent to Thomason and Kaufman’s level 1 on the borrowing scale (1988: 74ff), and is itself composed of two stages: at first it is non-basic vocabulary that is borrowed, and only later are more basic vocabulary items borrowed.
- **Stage Two** requires both a reasonable level of competence in the dominant language for the speakers doing the borrowing and a sufficient population of bilingual speakers. It is in this stage that discourse-pragmatic words are borrowed, most likely through code-mixing. These discourse-pragmatic words are used for the same functions in all the languages into which they are borrowed – to structure and comment on narrative, to perform social gestures, and to evaluate and express the speaker’s stance towards the

discourse. The use of these dominant-language discourse-pragmatic words is associated with a composite morphosyntax in the minority language, and it seems likely that much of this composite morphosyntax is motivated by the discourse-pragmatic words' retention of dominant-language grammatical requirements such as word order. These borrowing phenomena are roughly equivalent to those found in Thomason and Kaufman's level 2 on the borrowing scale (1988: 74ff).

- **Stage Three** assumes a level of speaker competence in the dominant language and bilingualism in the populace that is at least equal to, but most likely greater than, that found in Stage Two. It is in this stage that grammatical words without discourse-pragmatic functionality are borrowed: this can include an expansion in functionality of grammaticalized words borrowed in Stage Two as discourse-pragmatic words such that in Stage Three they are also used in their root meanings. For example, prepositions borrowed as hedges or evaluators of degree in Stage Two (both discourse-pragmatic functions) can also be used as spatial relators in Stage Three, and borrowed coordinators that are used only for narrative structure in Stage Two can also be used as grammatical coordinating conjunctions in Stage Three. In addition to the expansion in lexical borrowing, Stage Three is associated with further structural influence from the dominant language, including structural borrowing unrelated to lexical borrowing – the lexical and structural borrowings in Stage Three are roughly equivalent to those described in levels 3, 4 and 5 of the Thomason and Kaufman borrowing scale (1988: 74ff).

These three stages of lexical and structural borrowing will be revisited in the conclusion of this chapter, where the cross-linguistic data found in the case studies presented in sections 4.2, 4.3, and 4.4 will be integrated into the model.

It is necessary at this point to map out the relationship between synchronic styles of linguistic performance as described in Chapter Three and the diachronic stages of lexical and cultural borrowing described in the model above, as they are not exact parallels. The Tatar continuum of linguistic performance presented in Chapter Three is arranged according to language mixing, and an increase in the level of verbal hygiene leads to the increased de-Russification of Tatar, with “pure” Tatar at the extreme end of the continuum (but note that “pure” Tatar still contains Russian content words that have become conventional borrowings). Figure 14 below shows a more general representation of the “purification” of linguistic performance, where the vertical axis represents an increase in the level of verbal hygiene: it is only at the highest level of purification that dominant-language discourse-pragmatic words are not code-mixed.¹

Figure 14. Results of increase in level of verbal hygiene (synchronic style-shifting)

content words ²	system words	discourse-pragmatic words	–	↓ increase in level of verbal hygiene
	system words	discourse-pragmatic words	–	
		discourse-pragmatic words	–	
			no code-mixing	

The first row of the table is approximately what is produced in the two Tatar code-switching styles, although recall that these two styles cannot be mapped out with precision until they are recorded in future fieldwork. The third row of the table is the equivalent of Tatar-preferred style, and the fourth row is the equivalent of Tatar on-stage style. I have not heard Tatar linguistic performance that is equivalent to the second row, where both system words

¹ Self-reported introspection on the part of several bilingual speakers of English with whom I am acquainted echoes these findings, except for these speakers the discourse-pragmatic words from their first language (their preferred language) are what are removed last from their English speech (their second language).

² The content words in this column are not conventional borrowings.

and discourse-pragmatic words are code-mixed, but it is certainly theoretically possible, and may perhaps be found in style-shifting continua other than the one in the Tatar case.

By comparison, Figure 15 presents the diachronic stages of lexical infiltration described in the model laid out above; here the vertical axis is the passage of time.

Figure 15. Changes in composition of the borrowed lexicon (diachronic model)

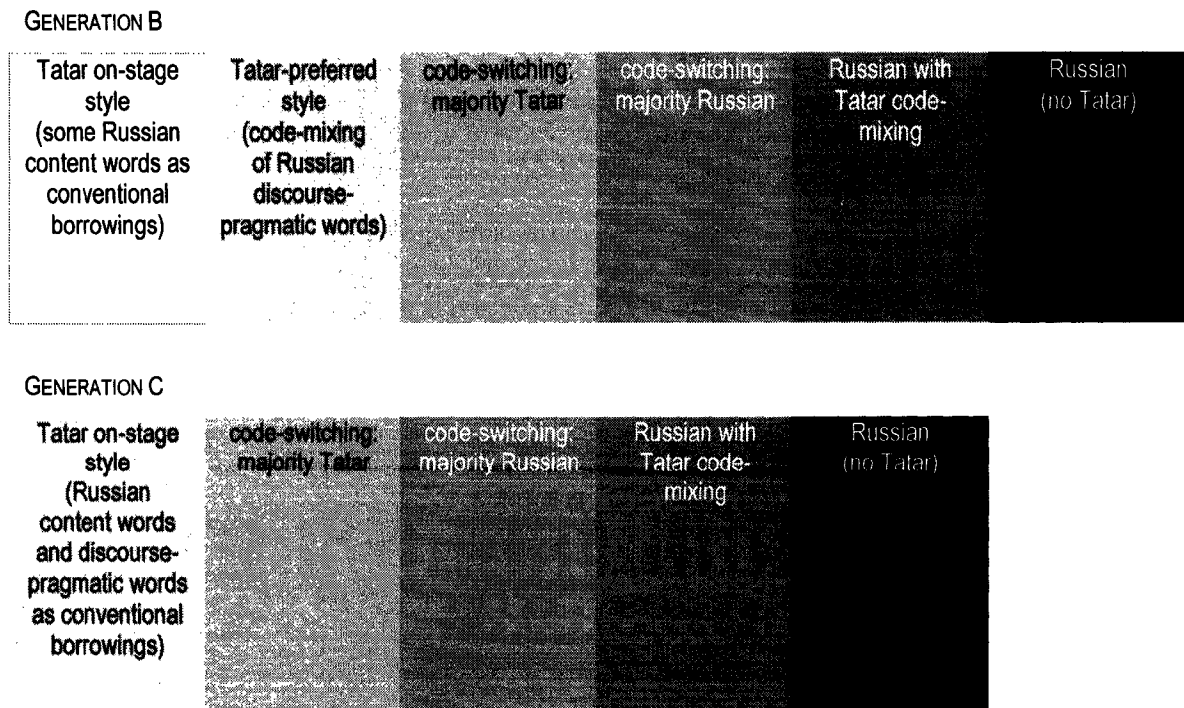
			no foreign words	<i>passage of time</i> ↓
		content words	–	
	discourse-pragmatic words	content words	–	
system words	discourse-pragmatic words	content words	–	

At the beginning stage (which can be called Stage Zero) there are no foreign words in the lexicon that are conventional borrowings. At some point during this stage, nonce borrowings – that is that is, code-mixed words – become conventional borrowings, and the language moves into the next stage, Stage One of the model presented here (the second row of the table). The third row of the table is Stage Two, and the fourth row is Stage Three. Presuming a continuum of language mixing like that of Tatar, with a pure version of the minority language at one end, a pure version of the dominant language at the other end, and various types of language mixing in between, suggests the following changes in available styles and language mixing over time:

- *Generation A.* Most pure version of minority language has no foreign elements as conventional borrowings (Stage Zero)
- *Generation B.* Most pure version of minority language has content words as conventional borrowings (Stage One)
- *Generation C.* Most pure version of minority language has content words and discourse-pragmatic words as conventional borrowings (Stage Two)
- *Generation D.* Most pure version of minority language has content words, discourse-pragmatic words, and system words/morphemes as conventional borrowings (Stage Three)
- (*Generation E.* Minority language speakers only produce code-switched styles)

I am using the term ‘generation’ quite loosely here, and it is meant to merely show the passage of time, not to suggest an age difference between speakers in one generation and speakers in another, nor to suggest the kinship relations between them; I imagine that in stable diglossic situations, for example, all three generations of a family (grandparents, parents, children) could be in Generation B of speakers. The young Tatar speakers cited in Chapter Three are Generation B, and their children (should they still speak Tatar at all) will perhaps be speakers in Generation C – today’s code-mixed Russian discourse-pragmatic words will in time most likely become conventional borrowings. Should this happen, the grammatical structure of Tatar on-stage style will change; today’s “pure” Tatar on-stage style, which has only conventional Russian content-word borrowings, will no longer exist, and what is today’s Tatar-preferred style will be tomorrow’s Tatar on-stage style. This potential reassignment of styles is shown in Figure 16:

Figure 16. Reassignment of styles by next Generation of speakers



In this form of language attrition the style that is the furthest away from Russian, on the Tatar end of the cline, will move closer and closer to Russian in successive generations, while pure Russian, on the other end of the cline, will remain stable.

Chapters Two and Three of this dissertation established that among Tatar speakers there is a prevalent “discourse of purity” where pure Tatar, which can alternatively be described as “de-Russified” Tatar, is an explicit ideal in both public and personal discourse. In the style of Tatar speech and writing where verbal hygiene is at its highest, the use of Russian words that are not conventional borrowings is quite rare, although code-mixed discourse-pragmatic words can be found. It is only when the level of verbal hygiene is lowered and discourse is more informal that regular and unconscious code-mixing of discourse-pragmatic words can be found, even among the most competent young Tatar speakers who are actively involved in the linguistic maintenance of their Tatar identity.

Present-day Tatar is by no means the only language where speakers who are striving for a “pure” ideal code-mix dominant-language words. For example, “indigenous purism and strict compartmentalization” are two characteristics of the language ideologies found among the Arizona Tewa, who are considered notable for their success in resisting cultural and linguistic assimilation (Kroskrity 1998). The emphasis on linguistic purism here is seen as stemming from performance registers associated with religion: “During ritual performance there is an explicit and enforced proscription against the use of foreign words and/or native vocabulary clearly identified with an...alien social dialect.” However, “the negative evaluation of instances of code-mixing in everyday speech by members of the Arizona Tewa speech community reflects not the prevalence of negative attitudes about these other languages, but rather the functioning of ceremonial speech as a local model of linguistic

prestige” (ibid. 107-108). This purism has kept the lexicon of Tewa relatively free from borrowings: there are less than twenty lexical borrowings after 100 years of contact with Apache, 150 years of contact with Spanish, and 191 years of contact with Hopi. However, although the “mixing of Tewa with either English or Hopi is explicitly devalued by members of the Tewa speech community” we find that “in unguarded speech some mixing does occur” (ibid.: 100), a situation parallel in some ways to the differentiation between Tatar on-stage style, which is essentially free of code-mixing, and Tatar-preferred style, in which code-mixing takes place. Kroskrity describes the structural borrowing phenomena in Tewa as “less salient” to speakers: among these borrowing phenomena we find a change in discourse structure, such that use of the evidential particle *ba* parallels that of the Hopi quotative particle *yaw* (ibid.: 110).

Another situation where speakers striving for a “pure” linguistic ideal allow dominant-language discourse-pragmatic words to “sneak through” is found in Irish radio performance as documented by Cotter (1996). Like Tatar, Irish is a contracting language, and while English is found in all functional domains in the Republic of Ireland, many of which are associated with power and prestige, Irish is found in very few functional domains. Unlike Tatar, there are comparatively few native speakers of Irish remaining, and the reclaiming of public functional domains such as Irish-language media is a conscious part of the cultural and linguistic revitalization process. Most of the Irish-language radio performers recorded by Cotter were not native speakers but instead learned Irish in school in programs that are also part of the revitalization process. What Cotter found is that radio personnel speaking on the air would use English discourse markers, words that are not standard borrowings in Irish, in otherwise Irish speech. Table 2 below shows the English discourse markers code-mixed in

Irish, along with their approximate Russian equivalents and an assessment of whether those Russian equivalents are found code-mixed in Tatar.

Table 2. English discourse markers found in Irish radio performance

English discourse marker	Approximate Russian equivalent	Found in Tatar?
well	<i>tak</i>	yes
yeah	<i>da</i>	yes
so	<i>tak; tak, chto</i>	yes
but	<i>a, no</i>	yes
because	<i>potomu chto</i>	yes
and	<i>i</i>	yes
now	<i>seychas, teper'</i>	no
y'know	none	n/a
I mean	none	n/a

Note that some of the Russian “equivalents” in the above table are not precise equivalents of their English counterparts; in particular, several of the Russian words have discourse-pragmatic function but are not “discourse markers” in Russian, that is, they do not function in Russian as their counterparts do in English. For example, Russian *potomu chto* is not used as an utterance-initial and periodically phonetically reduced discourse marker like English *because* – it is used only as a subordinator. And while Russian *tak* is used as a discourse particle that is essentially equivalent to English *so*, *tak chto* is not, and also is used only as a subordinator. A final difference is that while Cotter found that English *and* and *because* were infrequently code-mixed, the code-mixing of their Russian equivalents is quite common in Tatar-preferred style.

There are several other major differences between the code-mixing found in Irish and in Tatar. The English discourse-markers were code-mixed in a style where there is no code-mixing in Tatar: public performance of Tatar in such venues as radio and television, even when in an informal register, is usually in Tatar on-stage style, where the level of verbal hygiene is so high that code-mixing does not take place. Additionally, many of the speakers

recorded code-mixing English into Irish are not fully competent speakers, but rather heritage speakers or second-language learners, and apparently Irish-language teaching “falters at the level of discourse” (Cotter 1996: 148). The final difference in the two code-mixing phenomena is that Cotter found *only* discourse markers code-mixed into Irish, while Tatar speakers code-mix a range of discourse-pragmatic words, only a few of which are used as discourse markers in Russian. Even so, the parallels are significant: in both languages, when there is code-mixing in a style when speakers are striving for a pure version of the minority language, the only words code-mixed are discourse-pragmatic in function. Moreover, seven of the nine code-mixed English discourse words have Russian equivalents, and six out of seven of these Russian equivalents are found code-mixed in Tatar.

An examination of Hebrew code-mixing into English also shows interesting parallels. Olshtain and Blum-Kulka (1989), studying conversations of Israeli families for whom English is the home language, found an asymmetry of code-mixing: when English was the matrix language of conversation, there could be either code-switching into Hebrew or code-mixing of Hebrew words into otherwise English speech, but when Hebrew was the matrix language, there was only code-switching – code-mixing of English words into otherwise entirely Hebrew speech was not found. In other words, Olshtain and Blum-Kulka’s fieldwork gave them access to three mixed styles: English with Hebrew code-mixing (parallel to Tatar-preferred style), English with Hebrew code-switching, and Hebrew with English code-switching (parallel to the two Tatar-Russian code-switched styles). It does seem plausible, however, that other mixed styles exist, styles that were not appropriate for the context of their sociolinguistic study. In particular, it would be interesting to know if there is a Hebrew-English equivalent of Russian with Tatar code-mixing (described in section 3.2.3 of Chapter Three), that is, Hebrew discourse with occasional code-mixed English words.

Recall that this Tatar-Russian style is used by Tatar bilinguals in a self-conscious and deliberate way as both a marker of solidarity and as an invitation to switch conversational style towards the more Tatar end of the cline. If this style does exist in the repertoire of Israeli Hebrew-English bilinguals whose home language is English, perhaps it functions in the same way, which would explain why it would not be found at dinner-time conversations at home, where it would not be necessary to express ethnic solidarity nor to hint at language competence. In addition to the lack of complete stylistic data, another complication in drawing parallels is that the sociolinguistic context of English in Israel is unlike that of Irish in Ireland and Tatar in Tatarstan, most notably in that English is a requirement for many white collar jobs: “knowing English is a social asset to anyone living in Israel and English-speaking immigrants are strongly motivated to maintain their language, even when full acculturation in the new community is aspired to” (Olshtain and Blum-Kulka 1989: 62).

In contrast to both Cotter’s findings for Irish and my findings for Tatar, Olshtain and Blum-Kulka found that 64% of the code-mixed Hebrew words in their corpus were noun or verb phrases, while just 14% were what they label “discourse fillers” (ibid.: 68).³ The two discourse markers found in their article are *aval* ‘but, however’ and *davka* ‘specifically, peculiarly’, which is considered to be a Hebrew-specific discourse particle. Additional discourse-pragmatic Hebrew words can be found in Maschler (1994), where the data are drawn once again from the private conversations of Hebrew-English bilinguals living in Israel. The Hebrew discourse-pragmatic words found code-mixed in these two studies, along with their Russian equivalents and presence or absence in Tatar, are summarized in Table 3.

³ The other 22% of code-mixings fell into the following categories: Expressions of cultural content – 12.5%, Idiomatic expressions – 3%, and Others – 6.5%.

Table 3. Hebrew discourse markers found in Israeli English conversations

Hebrew discourse marker	Approximate Russian equivalent	Found in Tatar?
'aval 'but, however'	<i>no</i>	yes
'ax 'so'	<i>tak</i>	yes
'o 'or'	<i>ili</i>	yes
<i>bidyuk</i> 'exactly'	<i>tochno</i>	yes
<i>gam</i> 'also'	<i>tozbe</i>	yes
<i>ken</i> 'yes'	<i>da</i>	yes
<i>ki</i> 'because'	<i>potomu chto</i>	yes
<i>tishme'i</i> 'listen'	<i>slushai(te)</i>	yes
<i>ve</i> 'and'	<i>i</i>	yes
<i>vegam</i> 'and also'	<i>tozbe (?)</i>	yes (?)
'at <i>mevina</i> 'understand?'	<i>poniatno</i> (with rising intonation)	no
<i>beseder</i> 'all right'	<i>ladno</i>	no
<i>lebefex</i> 'on the contrary'	<i>naoborot</i>	no
<i>naxon</i> 'true, right'	<i>pravda</i>	no
<i>tir'i</i> 'look'	<i>smotri(te)</i>	no
'at <i>yoda</i> 'at 'y'know'	none	n/a
'axshav 'now'	none	n/a
<i>betax</i> 'sure'	none	n/a
<i>davka</i> 'specifically, peculiarly'	none	n/a

Here the overlap between Hebrew discourse-pragmatic words code-mixed in English and Russian discourse-pragmatic words code-mixed in Tatar is less complete than it was for Irish, but the parallels are still striking. There are fifteen Hebrew discourse-pragmatic words with approximate Russian equivalents, and ten out of these fifteen Russian words – or two-thirds – are found code-mixed in present-day Tatar. They include subordinators (Hebrew *ki* and Russian *potomu chto* 'because'), coordinators (Hebrew *ve* and Russian *i* 'and'), contrastors (Hebrew 'o and Russian *ili* 'or'), directives (Hebrew *tishme'i* and Russian *slushai(te)* 'listen'), and evaluators (Hebrew *bidyuk* and Russian *tochno* 'exactly').

Further parallels in discourse-pragmatic code-mixing are found closer to home (home for Tatars, that is) in nearby Uzbekistan. Uzbek belongs to a different branch of Turkic than Tatar – the Southeast branch – and Russian domination of Uzbekistan began significantly later, in the 19th century. Uzbekistan was a Union Republic during the Soviet

era, one level of political autonomy higher than Tatarstan, and is now an independent member of the CIS. As in Tatarstan, the post-Soviet Uzbek government has engaged in the promotion of the Uzbek language, and there appears to be a discourse of purity here as well, where “[s]ome Uzbeks bemoan the presence of Russian loans in their language and criticize code-switching, identifying these with the loss of Uzbek culture” (Baran 2000: 20). In addition, as in Tatarstan, Uzbekistan appears to be engaged in an officially supported purification of the lexicon, and “children reading the new Uzbek language textbooks discover that the Russian words used by their parents have been replaced with old Turkic, Arabic, or Persian words, or new Turkic coinages” (ibid. 31).

Baran (2000) presents a list of twelve Russian words and phrases that are found in what she calls “colloquial Uzbek” and an “informal register” but are not found in written Uzbek. These discourse-pragmatic words, which she refers to as “function words”, were present in the conversational speech of Uzbek-Russian bilinguals living in Tashkent, even when the speakers were Uzbek-dominant and “almost never” code-switched (ibid. 22), and are found in Table 4.

Table 4. Russian discourse-pragmatic words found in present-day Uzbek

Russian discourse-pragmatic word	Found in Tatar?
<i>i</i> ‘and’	yes
<i>ili</i> ‘or’	yes
<i>a</i> ‘and, but’	yes
<i>nu</i> ‘so, well’	yes
<i>naprimer</i> ‘for example’	yes
<i>prосто</i> ‘simply, just’	yes
<i>uzhe</i> ‘already’	yes
<i>srazu</i> ‘immediately’	yes
<i>da</i> ‘yes’	yes
<i>voobshche</i> ‘completely, in general’	yes
<i>tak, a tak</i> ‘like this, in general’	no(t yet)
<i>kaz raz</i> ‘just then’	no(t yet)

Baran believes that these discourse-pragmatic words have made the transition from “code-switched” words to borrowings that are in the “mental lexicon” of even Uzbek-dominant speakers (ibid. 21), but her data suggest that these Russian words are actually being code-mixed into otherwise Uzbek speech in a style that appears to be the equivalent of Tatar-preferred style. As in Tatar-preferred style, the code-mixed Russian words retain their Russian grammatical requirements and the discourse produced is a composite of Uzbek and Russian grammar (in Baran’s analysis, “the structures permitted by *a* and *uzbe* [two of the words listed above] are ungrammatical when unmixed Uzbek is used” (ibid. 23)). Nearly all of the Russian discourse-pragmatic words listed by Baran as code-mixed in Uzbek are found code-mixed in Tatar and are presented in Chapter Three (and note that Russian *a* ‘however,’ is already a conventional borrowing in Tatar, found in both oral and written registers). The two exceptions, *tak/a tak* ‘like this, in general’ and *kak raz* ‘just then,’ may not actually be exceptions – these are phrases that I believe that I have heard code-mixed into Tatar at times when I did not have note-taking materials available. Therefore, rather than making a definitive claim one way or the other, I have simply labeled their “found in Tatar?” status as “not yet.” Finally, note that while there are 61 code-mixed Russian discourse-pragmatic words found in my Tatar corpus, Baran presents just 12 in Uzbek; it would be interesting to know if there are further parallels in the code-mixing of Russian words in Uzbek and Tatar that were simply not presented in her brief article due to space limitations.

Hypotheses for the motivations and mechanisms of discourse-pragmatic code-mixing are found in Chapter Five. The next three sections, 4.2, 4.3, and 4.4 are case studies of the discourse-pragmatic borrowings and related borrowing phenomena found in three different language contact situations and examine: (1) the influence of Russian and Polish on Karaim; (2) the influence on Tatar of Arabic and Persian, two languages that served as

superstrates for nearly a millenium; and (3) the influence of Turkish on Asia Minor Greek. These three contact situations were chosen because of their relevance to present-day Tatar and the code-mixing of discourse-pragmatic words as described in Chapter Three. All involve Turkic languages: for one contact situation, the Turkic language is the dominant language doing the influencing; for the other two, it is the Turkic language that is being influenced. The three case studies work together to show that not only is discourse-pragmatic functionality key to understanding both lexical and structural borrowing, but that the same discourse pragmatic words are borrowed from the same dominant language in separate contact situations, and when there are no contact languages in common it is words with comparable discourse-pragmatic functionality that are borrowed again and again, and used with the same function, frequency, and placement.

4.2 Case Study One: The influence of Slavic on Karaim

Karaim, like Tatar, is a Turkic language of the Kipchak branch that has been in contact with Slavic languages for hundreds of years. The Karaim are sectarian Jews who had migrated from the Russian steppes to Crimea by the 13th century (Schur 1995) and from Crimea were invited to settle in Lithuania in the late 14th century (Zajaczkowski 1961). Karaim settlements were typically quite small, numbering at most hundreds of people (Schur 1995), and Karaim “has always been a sociolinguistically weak code dominated by the regional standard languages, Polish, Russian and Lithuanian, and their non-standard local variants” and “exhibits a high degree of convergence with the areally dominant codes” (Csató, in press: 1). The Karaim have been undergoing multi-generational language shift for some time; there are

currently approximately 40 speakers of the Lithuanian dialect and six speakers of the Ukrainian dialect (Csató 1999b: 1).

Another similarity between Karaim and Tatar is a language ideology of purism. An early 20th century language development and standardization movement, permanently interrupted by the outbreak of the second world war, had purist elements; language reformers explicitly promoted words of native stock over borrowings in their attempts to develop a standard Karaim literary language (Csató 1999a: 84). Most of the remaining speakers of Karaim belong to the pre-war generation, and thus were influenced by these purist language ideologies, where “pure” Karaim – Karaim without lexical borrowings – is an explicit ideal. Linguistic purism in Tatar can cause communicative issues for less competent speakers who feel compelled to speak in the de-Russified Tatar on-stage style and thus produce halting, ungrammatical speech; they prioritize ideology over communication. Purism appears to have a similar effect among the Karaim. “In communicative situations where they feel ‘observed,’ the conversation in Karaim often breaks down after some cliché of greeting and asking about the other’s health because the speakers are afraid of using the language ‘incorrectly’ ” (ibid. 85).

Despite this ideology of purity, speakers of Karaim, which is much further along in the process of language shift than Tatar and has been more heavily influenced by its contact languages, rely heavily on both lexical borrowings and upon code-switching. There are apparently descriptive rules for code-switching that are perceptible, if not articulated: “A Karaim can tell immediately whether a person uses the foreign elements in the real Karaim way or just ‘mixes’ foreign words with the Karaim language” (ibid. 88). Additionally, speakers who actualize the ideal of “pure” Karaim mark themselves as non-native, and not entirely competent in the communicative norms for Karaim. One Karaim woman, in

reference to a Turcologist who speaks “pure, Turkic” Karaim, stated that “his way of speaking is not *b’iz’im karaycha* ‘our Karaim’ ” (ibid. 88). Karaim speakers of the late 20th century appear to be in Generation D in the typology laid out in section 4.1 above, that is, the style of Karaim produced for the fieldworker in interviews (presumably a more “performance” style, as discussed in section 1.4 of Chapter One), has dominant-language content words, discourse-pragmatic words, and system words and morphemes.

The foreign discourse-pragmatic words found in Karaim appear to be standard borrowings, but given both the similarities in circumstances between Karaim and Tatar and the similarities in the placement and frequency of these discourse-pragmatic words, it seems quite likely that code-mixing is the mechanism by which they were borrowed. Karaim has borrowed the following discourse-pragmatic words from Slavic (Russian and Polish):

[4.1]

- *a* ‘and’
- *i* ‘and’
- *to* ‘then’
- *nu* ‘well’
- *nu to* ‘well then’
- *vot* ‘there’
- *móc* ‘to be able’
- *uzhe* ‘already’
- *okolo* ‘around’

With the exception of *to* ‘then’, all of the Slavic discourse-pragmatic borrowings found in Karaim are words that are code-mixed in Tatar. Additionally, it is entirely possible that there is more overlap of borrowed or code-mixed Russian discourse-pragmatic words – the corpus of Karaim data to which I had access is far from exhaustive. The above words were found in example sentences in various articles by Éva Ágnes Csató, who, while writing mostly about Slavic-Karaim contact phenomena, was not writing specifically on lexical borrowing. It is

entirely possible that a more extensive set of data would yield additional Slavic discourse-pragmatic words common to both Tatar and Karaim; in Tatar as code-mixed words in the process of being borrowed, and in Karaim as now-conventional borrowings.

Karaim speakers use two Slavic conjunctions, *a* and *i*, as found in examples [4.2] and [4.3]. All borrowings from Slavic, both discourse-pragmatic and content words, are presented in boldface in this section:

[4.2] Karaim borrowing of Slavic *a*

A	ɣal'ə	n'e	koduy	ish'l'-ey-s'?	M'en'	Diana-bə
and	now	what	you	do-A.NONPST-2PS	I	Diana-WITH
	yaz-a-m		b'it'ik.			
	write-A.NONPST-1PS		letter.			

'And what are you doing now? I am writing a letter together with Diana.' (Csató 2000a: 680)

Recall that Russian *a* is also now a standard borrowing in Tatar, one that was borrowed at an earlier stage and phonetically altered to *ä*. As with all of the Russian words code-mixed in Tatar, all of the Slavic borrowings in Karaim appear to remain true to Slavic pronunciation. Here in Karaim, as in Russian, *a* is used to introduce a new topical unit (Grenoble 1999: 180), and is thus found in sentence-initial position (it can also be found clause-initially in the second clause in a sentence).

The Slavic coordinating conjunction *i*, along with several other Slavic borrowings, is shown in example [4.3].

[4.3] Karaim borrowing of *i*

I **uzhe** **bu** **fotograf** **tur-at** **kolo** Bas'ya-nin.
and **already** **this** **photographer** stand-A.NONPST-3PS **by** Bas'ya-GEN
'And this **photographer** is **already** standing **near** Basia.' (Csató 2000b: 269)

Here we see not only a high frequency of words of Slavic origin, but also that three of the four Slavic borrowings in this utterance are non-content words – Russian *uzhe* 'already' and Polish *kolo* 'near' will be discussed later in this section. The discourse-pragmatic function of Russian *i* 'and' is consistent in Karaim, Tatar, and Russian, and parallel to that of English *and* – it coordinates idea units and continues a speaker's action (Schiffrin 1987: 125).

The remaining Slavic discourse-pragmatic borrowings into Karaim that are words used for metacommentary and deixis, or "information packaging" (Grenoble 1998). The first of these is the Russian temporal adverb *to*, usually glossed in this meaning as 'then' and described by Csató as a discourse operator used to "mark a vague consecutive relation" (1999b: 59). Here it can be seen as loosely parallel to the English temporal adverb *then*, which when used as a discourse marker marks succession in discourse time (Schiffrin 1987: 250ff). Note that *to* is the only Slavic borrowing in Karaim not found in my Tatar corpus, but I believe that its absence should be interpreted less as an indication that *to* is not code-mixed in Tatar-preferred style, and more as an indication that the corpus is not yet complete. Indeed, there are related words and phrases using *to* that are found code-mixed in Tatar-preferred style: *a to* 'else', which is the combination of *a* 'and/but' and *to* 'then' (and thus can be more literally translated as 'but then'), and *zato* 'but on the other hand', which is a compound word consisting of the preposition *za* 'for' and *to* 'then'. I believe that *to* is most likely code-mixed in Tatar to structure discourse, just as it is in Karaim. This discourse-

structuring function in Karaim can be seen in example [4.4], where *to* is the only Slavic borrowing:

[4.4] Karaim borrowing of *to*

Son n'echik chikh-sa **to** k'er'äk alma n'in'd'ikol'ek savut.
end when emerge-COND **then** need take-INF some pot
'After they have emerged, **then** you need to prepare the pot.' (Csató 1999b: 59)

This use of *to* is interpreted by Csató as a “specific strategy applied to mark a constituent as pragmatically prominent” and a construction found in a large linguistic area including Slavonic languages, Turkic languages in contact with Slavonic, Hungarian, and Modern Greek, among others (1999b: 59). Although Csató does not give an alternative way of expressing this utterance using only native Karaim means of subordination (perhaps because there no longer is one), this use of *to* as shown here is most likely associated with a composite morphosyntax – temporal relations in other Kipchak languages are usually expressed synthetically by means of verbal suffixes, but in example [4.4] the temporal relation is expressed with a combination of Karaim verbal forms and a Slavic temporal marker.

In addition to being used independently, *to* ‘then’ is also used in combination with *nu* ‘well, so’ in order to structure discourse, as seen in example [4.5]:

[4.5] Karaim borrowing of Slavic *nu to*

Vot	ɣal'ə	V'il'n'ius'-d'ä	bun'd'i	k'en'esa	k'ör'k'-l'ü	nu	to
look	now	Vilnius-LOC	such	kenesa	beautiful	now	this
tuy-du-m	<i>k'i</i>	ayt-ad-lar		<i>k'i</i>	bar-ir-lar		ɣal'ə
hear-PST-1PS	<i>that</i>	say:A.NONPST-3PP		<i>that</i>	go-R.NONPST-3PP		now
ulan-lar-ɣa	ulan-lar da	anda	ür'an-'ir-l'är		karaychə		s'öz'la-m'ä.
child-PL-DAT	child-PL also	there	learn-R.NONPST-3PP		Karaim		speak-INF.

'**Look**, this *kenesa* in Vilnius is so beautiful now, **and** I have heard *that* they say *that* they will now go to the children and the children will learn to speak Karaim there.' (Csató 2000a: 678)

Once again, the only Slavic borrowings in the utterance are discourse-pragmatic terms: *nu to*, used to introduce the second proposition of the sentence, and Russian *vot*, used as a forward-looking directive to draw attention to the statement about to be made – just as it is used in Russian, and just as it is used when code-mixed in Tatar (as seen in section 3.4.1d in Chapter 3). While Csató translates *nu to* as 'now this' in her interlinear gloss, it could just as easily be glossed as 'well then,' and a more literal free translation of the sentence would be something like, "Look, now in Vilnius is such a beautiful *kenesa*, **well then** I heard *that* they say *that* they will now go to the children, and the children will learn to speak Karaim there." Note that the Karaim word *k'i*, along with its gloss 'that', is italicized in the example and translation above; this *k'i* is the same Persian *ki* found in Tatar, which will be discussed in section 4.3.2 of this chapter. Not only are all of the borrowings in this sentence discourse-pragmatic words, but one of them is from an earlier stage of borrowing from a different language.

Russian *vot* can also be found in Karaim with a usage much like the English discourse marker *well*, as seen in example [4.6]:

[4.6] Karaim borrowing of Russian *voť*

Vot buley *vakht* ash-a-t.
well that *time* go-A.NONPST-3PS
 ‘Well, that’s how *time* goes’ (Csató 1998: 49)

Note that *vakht* ‘time’ is a borrowing from Arabic, parallel to *wakiť* in Tatar and *vakit* in Turkish. Russian *nu* can also be translated as ‘well’ and can be used similarly, as found in example [4.7]:

[4.7] Karaim borrowing of Russian *nu*

Nu	n’ech’ikb’ir	bala	oshukac’	et’-t’i	b’ir	tashcha-ni...
well	how a	young man	cheat	do-PAST	one	girl-ACC
Nu	da	ganglakh	uzhe	spov’edac’	et’-at’	alar-ni,
well	and	Catholic priest	already	hear confession	do-A.NONPST	they-ACC
<i>k’i</i>	pov’in’n’i	e-d’-lar	buley	da	buley...	
<i>that</i>	ought to	be-PAST-3PP	so	and	so	

‘Well, how a young man **gets** a girl **into trouble**, and the Catholic priest **hears** their **confession** so *that* they **ought to** do so and so.’ (Csató 2000b: 267)

Here *nu* is used twice: once to introduce a topic and once to signal its continuation; both of these discourse functions are found when *nu* is used in Russian (Grenoble 1998: 181). There are three kinds of borrowings found in this example. From Persian we once again see *k’i* ‘that’, used as a standard subordinating conjunction. From Russian we have *nu* ‘well’ as a discourse marker, and *uzhe* ‘already’, which here seems to be used less as an evaluative and more as an intensifier. Finally, from Polish we have verbs in the infinitive, borrowed in the standard Turkic manner with the auxiliary verb *et’* ‘do.’

Two further examples, [4.8] and [4.9], show that this pattern of borrowing – where the only Slavic borrowings are either prototypical content words (nouns and occasionally verbs) or discourse pragmatic words – is the standard in present-day Karaim:⁴

[4.8] Karaim borrowing of Russian *nu to* and nouns

Nu to bashla-di ish'l'ä-m'ä anar bu **d'ikhani'ye**
well then begin-PAST do-INF she-DAT this **respiration**

yattir'iz-di yer üs'n'ü, **kukhn'a-dä.**
 lay down-PAST floor on **kitchen-LOC.**

'**Well, then** he started to give her **artificial respiration**, he laid her down on the floor, in the **kitchen.**' (Csató 2000b: 266)

Example [4.9] is from later in the same narrative:

[4.9] Karaim borrowing of Russian *to*, verbs, and noun

Vizvat' et'-t'i-m **skora-ni,** k'el'-m'i-t'.
call do-PAST-1PS **ambulance-ACC** come-NEG-Ä.NONPST.

L'il'ä-g'ä **pazvont** et'-t'i-m, **to skoray** k'el'-d'i.
 Lilia-DAT **call** do-PAST-1PS **then ambulance** come-PAST.

'I **called** the **ambulance**, it doesn't come. I **called** Lilia **and** the **ambulance** came.'⁴ (Csató 2000b: 267)

⁴ This is the precise pattern of Slavic code-mixing and borrowings found in Tatar-preferred style.

Note that in these two examples Russian nouns are treated like Karaim nouns in that they take Karaim morphology: *kukbn'a* 'kitchen' is in the locative case, and *skora* 'ambulance' is in the definite accusative case. In example [4.8], Russian *nu to* is used to mark a continuation in the discourse and, like *to* in example [4.9], provides narrative structure and event sequencing.

The remaining three discourse-pragmatic words borrowed from Slavic into Karaim are evaluative in function: *móc* 'to be able' is used as an epistemic verb, *uzbe* is used as an evaluator of time, and *okolo* is used as a hedge. The Karaim borrowing of Polish *móc* 'to be able' differs both from the Tatar borrowing of Russian *moch'* 'to be able' and from the Karaim borrowing of other Polish verbs. Here is an example of Polish *móc*:

[4.10] Karaim borrowing of Polish *móc* 'to be able'

Gal'ə t'ir'il'-ä-m **mog-ə-m** ayt-ma k'i t'ir'l'ik khor t'üv'ül'.
 Now live-A.NONPST-1PS **can**-A.NONPST-1PS say-INF that life bad not.
 "Now I live and I **can** say that life is not bad." (Csató 2000a: 687)

When Tatar speakers code-mix Russian *moch'*, it is only in the third-person singular form *mozhet*. And other Polish verbs are borrowed into Karaim in the noun-like infinitive form with the borrowing mechanism of being verbalized with the auxiliary verb *et'* 'to do.' However, as can be seen here, Polish *móc* is neither used in a fixed third-person singular form, nor borrowed in the infinitive and treated like a noun, but rather fully nativized and used like a Karaim verbal stem. The form of the verb used as the stem here is in fact the form of the stem for the first-person singular present in Polish: *móc* is a verb that has a morphophonological alternation such that the final consonant of the first-person singular and third-person plural verb stem is [g], while all other persons have a verb stem that ends in [3]. Therefore, what we seem to have here is a borrowing that is sensitive to Polish grammar,

where the verb stem for the first-person singular in Polish is the one used as the verb stem for a first-person singular in Karaim.

Móc seems to be the only Polish verb for which this is the case. Additionally, all of the other modals used in Karaim – at least those found in the data available – are natively Turkic: the use of the question particle to question possibility; the use of the verb *bolal-* to express permission and ability; and the use of the verb *yara-* to express permission or objective possibility. Russian *moch'* is code-mixed into Tatar as the fixed third-person singular *mozhet* and used as an epistemic meaning ‘might, maybe’: in both Karaim and Tatar, *moch'/móc* appears to be the only dominant-language modal verb that is either borrowed or code-mixed.

The evaluative Russian *uzhe* ‘already’ is “often used to stress present relevance” in Karaim (Csató 2000: 681), as can be seen in example [4.11]:

[4.11] Karaim borrowing of Russian *uzhe*

Gal'ə	maya	uzhe	yet'm'ish	yil	tol-du.
Now	I-DAT	already	seventy	year	fulfill-PAST.

“I have now completed my 70th year.” (Csató 2000a: 681)

Note that *uzhe* appears to be moving towards a semantic shift as a general intensifier, as can be seen in example [4.7] above, where it does not seem to have a real temporal meaning – here is an excerpt from that example:

[4.12] Russian *uzhe* as intensifier in Karaim

Nu da ganglakh **uzhe** **spov'edac'** et'-ät' alar-ni,
well and Catholic priest **already** **hear confession** do-A.NONPST they-ACC

k'i **pov'in'n'i** e-d'lär buley da buley...
that **ought to** be-PAST-3PP so and so

'...and the Catholic priest **hears** their **confession** so *that* they **ought to** do so and so.'
 (Csató 2000b: 267)

While Csátó has translated *uzhe* as 'already' in the interlinear gloss, she does not include 'already' in the free translation, because this *uzhe* is not functioning as an evaluator of time. In Russian, the emphatic particle *uzh*, often used as a general intensifier, is a reduced version of *uzhe* – this semantic shift appears to be taking place once more in Karaim, albeit without the phonetic reduction. Tatar *inde* is also polysemous, generally translated with either Russian *uzhe* 'already' or *zhe*, an emphatic particle, although it is not clear if this semantic shift was independent or externally motivated.

The final evaluative discourse-pragmatic word borrowed into Karaim from Russian is *okolo* 'around, about', used in Karaim precisely as it is in Tatar:

[4.13] Karaim borrowing of Russian *okolo*

Okolo b'ir afta-nin Trox-ta kal-i-m.
about one week-GEN Trakai-LOC stay-AOR-1PS
 'I will stay **about** a week in Trakai.' (Csátó 1999b: 55)

Here, as in Tatar, Russian *okolo* in Karaim is used in its grammaticalized, evaluative meaning as a hedge, rather than in its root spatial meaning to mean literally 'around' or 'about'. Additionally, its use here is accompanied by a composite morphosyntax: *okolo*, which is a preposition in Russian, retains its prepositional nature and precedes the head it is modifying; by contrast, Karaim, like other Kipchak languages, is traditionally postpositional. Moreover,

okolo retains its Russian grammatical requirements; the modified noun *afta* ‘week’ head is in the genitive (with Karaim morphology), which is the case required by this preposition in Russian.

We have seen that eight of the nine Slavic discourse-pragmatic words borrowed into Karaim are also found code-mixed in Tatar preferred-style. Additionally – with the exception of *móc* ‘to be able’, borrowed into Karaim as a verb stem and inflected for person with native morphology, but code-mixed in Tatar only in the fixed third-person singular form – these Slavic discourse-pragmatic words are used identically in Karaim and in Tatar. Their frequency and placement within utterances is essentially the same, and if one were to compare just the translations, where Slavic discourse-pragmatic words are marked in boldface, one would be unable to distinguish which translation was for Karaim and which translation was for Tatar.

However, Karaim is at a much later stage of language shift and borrowing than Tatar; not only is it moribund (the current generation of speakers is most likely the last generation), but it has also been more influenced by the locally dominant languages than Tatar has been influenced by Russian. In addition to the discourse-pragmatic borrowings described above, there are other borrowings and elements of structural influence that show Karaim to be in the third stage of lexical borrowing described in this model, and further along the Thomason and Kaufman borrowing scale (1988: 74ff). Because both Karaim and Tatar are members of the Kipchak branch of Turkish (itself a relatively shallow family) and the dominant languages in both situations are comparable (Russian and Russian/Polish), the Karaim structural innovations can be seen as potential future innovations for Tatar. In other words, borrowing phenomena found in Karaim today may possibly be found in the Tatar of tomorrow.

While the only Russian non-content words code-mixed into Tatar by fully competent young speakers are discourse-pragmatic in function, Karaim has borrowed grammatical non-content words without metalinguistic functionality. For example, Russian *okolo* ‘around, about’, seen above in example [4.13], is used metalinguistically as a hedge. However, Polish *kolo* is used in Karaim in its root spatial meaning, as seen in example [4.3] above, shown again here:

[4.14] Karaim borrowing of Polish *kolo*

I **uzhe** **bu** **fotograf** **tur-at** **kolo** **Bas'ya-nin.**
and **already** **this** **photographer** **stand-A.NONPST-3PS** **by** **Bas'ya-GEN**
‘**And this photographer is already standing near Basia.**’ (Csató 2000b: 269)

While the two Russian non-content words used in this sentence, *i* ‘and’ and *uzhe* ‘already’, have metalinguistic function, the Polish preposition is used as a spatial relator, and the object of the preposition is produced in the genitive, just as it would be in Polish. However, the genitive in this case is produced with Karaim morphology (just as it was with Russian *okolo* in example [4.13] above), thus producing a phrase – *kolo Bas'yanin* ‘by Basia’ – that is a morphosyntactic composite. It is not clear if Russian *okolo* is also used in Karaim in its spatial meaning – due to the paucity of data, it is difficult to make an absolute claim that it is not.

Polish *kolo* is not the only Slavic preposition used in Karaim – also borrowed are Russian *s* ‘from’ and *za* ‘for’. Karaim use of Russian *s* is shown in example [4.15]:

[4.15] Karaim borrowing of Russian *с*

Trox-ta	anuz	yokh	e-d'i	gramafon,	a	ata-m
Trakai-LOC	yet	not existing	be-past	gramophone	and/but	father-POSS.1PS
k'el't'ir-d'i	s	Petrograd-a	gramafon.			
bring-PST	from	Petrograd-GEN	gramophone.			

'There was no **gramophone** in Trakai, **but** my father brought a **gramophone from** Petersburg.' (Csató 2000b: 269)

Once again, the preposition *с* precedes the noun it is modifying, and once again the Slavic preposition retains its grammatical requirements, here that the modified head be in the genitive case. However, the genitive here is not produced with Karaim morphology (*-nIn*, as seen above in example [4.13]), but with Slavic *-a*, the genitive suffix for masculine and neuter nouns, and thus appropriate for masculine *Petrograd*. It is not clear if there is free variation between the use of the Karaim genitive suffix and the Slavic genitive suffix with borrowed Slavic prepositions, or if some Slavic prepositions somehow trigger the use of Slavic morphology while others do not, but what is clear is that this represents a further penetration of Slavic into Karaim grammar.

Russian *за*, which has a root meaning of 'behind' and is polysemous, with grammaticalized meanings of 'for' and 'after', *inter alia*, is found in a calque of the Slavic idiom used to mean "What kind of ... is it?":

[4.14] Karaim borrowing of Russian *за*

N'e	bu	za	karay?
what	this	for	Karaim

'What kind of Karaim is this?' (Csató 2000b: 269)

The Russian equivalent, which is the model for the calque, is as follows:

[4.15] Russian equivalent

Chto eto za X?
what this for X
'What kind of X is this?'

Further syntactic influence can be found not in the borrowing of Slavic prepositions, but in the calquing of Slavic syntax. For example, Karaim, like most Turkic languages, does not have an instrumental case, while Slavic does – under the influence of Slavic, the clitic postposition *bə* has expanded from its comitative function to be used as a kind of 'case marking' that is parallel to Slavic instrumental (Csató, in press: 3), as can be seen in example [4.16]:

[4.16] Karaim *bə* used like Slavic instrumental

Ol e-d'i khazzan-bə
He be-PAST hazzan-WITH
'He was hazzan (leader of the Karaim community).' (Csató, in press: 3)

The Russian equivalent is as follows:

[4.17] Russian equivalent with instrumental case

On byl khazzan-om.
He be-PAST khazzan-INST
'He was hazzan.'

In Russian, the past tense of the verb *byl'* 'to be' requires a predicate noun in instrumental case, and it is this requirement that is being paralleled in Karaim. The non-final position of the verb combines with the new use of the enclitic *bə* 'with' to create an utterance that is both morphologically and syntactically parallel to Russian. The Tatar cognate of Karaim *bə* is the comitative *belän*, which is used as a postposition. While Tatar *belän* 'with' is periodically

used in calques from Russian, it is as the equivalent of the Russian preposition *s* ‘with’, and not to create an instrumental-like case in instances where that case is found in Russian without the prepositions. Additionally, this calqued use of *belän* seems to be limited to a few set expressions, for example, wishes for a happy holiday or birthday, as seen in example [4.18]:

[4.18] Tatar *belän* ‘with’ as a calque of Russian *s*

a. Tatar			Russian		
Tu-gan	kön-egez	belän	S	dnyom	rozhdeni-ya
be.born-PAST.PART	day- POSS.2PP	with	with	day.INST	birth-GEN
‘Happy birthday!’			‘Happy birthday!’		
b.					
Bäyräm	belän		S	prazdnik-om	
holiday	with		with	holiday-INST	
‘Happy holiday!’			‘Happy holiday!’		

Tatar *belän* is otherwise not generally used as an equivalent of Russian comitative *s*. Additionally, the two expressions shown in example [4.18] are readily perceived as calques by many Tatars, some of whom attempt to use alternate, non-calqued, expressions.

Another productive morphosyntactic borrowing from Slavic into Karaim is a pattern for deriving compound nouns. The Turkic morphological process used to create compound nouns is known as *izafet* and consists of a first, unmarked, noun and a second noun that immediately follows it and is marked with a possessive suffix that is identical to the third-person singular possessive *-(s)I*. This *izafet* construction can be found in many of the Tatar examples in Chapter Three, such as those repeated in [4.19]:

[4.19] The Tatar *izafet* construction

uzeshchen-när konsert-i,
amateur-PL concert-POSS
'amateur concert' (example 3.14)

Mishär kiz-lar-i
Mishar girl-PL-POSS
'Mishar girls' (example 3.17)

Khäter Kön-e
memory day-POSS
'Memory Day' (example 3.23)

Karaim also has the *izafet* construction, but has innovated a new construction for compound nouns, one that is parallel to the Slavic model of Noun + Noun-PL.GEN, as can be seen in example [4.20]:

[4.20] Karaim innovated construction for compound nouns

sawuxturuwchu t'ish'-lär'-n'in'
doctor tooth-PL-GEN
'doctor of teeth', i.e. 'dentist' (Csató in press: 4)

A typical Polish equivalent would be:

[4.21] Polish compound noun

lekarz chorób dziecięcych
doctor illness-PL.GEN child-PL.GEN
'doctor of child illness', i.e., 'pediatrician' (Csató in press: 4)

This innovated Karaim construction, clearly based on the Slavic model, is used productively by speakers (Csató in press: 5). Tatar, by contrast, has no such construction. What one will find is the use of the Russian derivational morpheme *-skii* to create an adjective that then modifies a head noun, for example, *Bolgarskii stil* 'Bolgar style' as opposed to *Bolgar stil-e*

‘Bolgar style’. However, this construction, as discussed in section 3.5 of Chapter Three, is neither conventionalized nor frequently heard, and is associated with less-competent Tatar speakers.

The Karaim enclitic *da* ‘and’, cognate with the Tatar enclitic *da* ‘and, also’ (which assimilates in consonant voicing and vowel quality to the preceding word), also seems to have an innovative usage that is based on a Slavic model. In Tatar, *da* always follows the items or propositions that are being conjoined, and is usually found after each item or proposition being linked. Here is an excerpt from example [3.11] in Chapter Three:

[4.22] Tatar use of the enclitic *da*

R.-nĭ da chakĭr klub-ka.
R-ACC also invite-IMP club-DAT
‘Invite R. as well to the club.’

Tatar speakers sometimes code-mix Russian *tozhe* ‘also, as well’ in lieu of *da*, as can be seen in the sentence following the one in example [4.22], also taken from example [7] in Chapter Three:

[4.23] Tatar use of Russian *tozhe* in lieu of *da*

Anga tozhe ber kiz tab-ar-sing
him-DAT also a girl find-AOR-2PS
‘You will **find** a girl for him as well.’

While Karaim does not appear to have borrowed *tozhe* ‘also, as well’ from Russian, it does seem to have innovated a use of *da* that is based on Slavic *i* ‘and’ and *tozhe* ‘also’, as seen in example [4.24]:

[4.24] Karaim *da* modeled on Russian *i*

Da ür'ät'üv'ch'ü-l'är aytİR e-d'-l'är n'in'd'i yakhshİ karay-lar.
 and teacher-PL say:R.NONPST be-PST-3PP what good Karaim-PL.
 'And the teachers used to say that the Karaims are so good.' (Csató 2000a: 682)

In Tatar, *da*, as an enclitic, can never be in sentence-initial position – the discourse-pragmatic word used sentence-initially in standard literary Tatar to introduce a conjoined proposition is the Persian loanword *häm* 'and', while in Tatar-preferred style Russian *i* 'and' is sometimes used. This Karaim use of *da* is parallel to the use of borrowed Russian *i* in Karaim, shown here once again in example [4.25]:

[4.25] Karaim borrowing of *i*

I uzhe bu fotograf tur-at kolo Bas'ya-nİN.
and already this photographer stand-A.NONPST-3PS by Bas'ya-GEN
 'And this photographer is already standing near Basia.' (Csató 2000b: 269)

In present-day Tatar, the syntactic functionality of the enclitic *da* has remained distinct from its non-native equivalents *häm* (conventional), *tozhe* (non-standard), and *i* (non-standard), but in Karaim, *da* seems to have changed from a characteristically Turkic enclitic to a coordinating conjunction that is used exactly as *i* is used in Russian.

One final Karaim innovation, found in the syntactic realization of relativization, further demonstrates the level of Russian structural influence. As shown in Chapter Three, the use of code-mixed Russian coordinators in Tatar is associated with a composite morphosyntax – whereas in an entirely Tatar utterance, subordinate clauses are embedded pre-head and subordinated using participles, nominalizations, and gerunds, a code-mixed Russian coordinator will introduce a subordinate clause that is not embedded, as would be standard, but rather follows the main clause. In Karaim, this composite morphosyntax can

be found without the use of a Slavic borrowing, as shown in example [4.26], which shows a native Karaim question word used as a Slavic-style relativizer:

[4.26] Karaim use of *kayda* ‘where’ as a relativizer

E-d’i-k Yusuf-ta, yilisuv-da kayda astriï yakhshi yuvu-n-du-kh.
be-PAST-1PP Yusuf-LOC sauna-LOC where very good wash-RFL-PAST-1PP.
‘We were at Yusuf’s (house), in the sauna, where we had such a nice bath.’
(Csató 2000a: 678)

Compare the syntax of the Karaim sentence with its equivalent in Russian:

[4.27] Russian equivalent

My byl-i u Yusuf-a, v saune, gde my ochen’ khorosho
we be-PAST.PL at Yusuf-GEN in sauna.LOC where we very well
pomyl-i-s’.
wash.PERF-PAST.PL-RFL

‘We were at Yusuf’s (house), in the sauna, where we had such a nice bath.’

The Karaim sentence has no pronouns (required in Russian because the past tense marks number but not person), but otherwise it is a word-for-word equivalent of the Russian – *kayda* ‘where’ has expanded in scope so that now, like Russian *gde* ‘where’, it can be used not only as a question word but also as a relative pronoun, what Myers-Scotton calls a convergence of “lexico-semantic conceptual structure” (Myers-Scotton 1992). Tatar *kayda* ‘where’ is not yet used as a relative pronoun in standard speech, and while it can occasionally be found as a relativizer in more casual styles of speech, this use of *kayda* is apparently immediately recognizable to competent speakers as a calque and disdained by them as not “real” Tatar (Amil Nur, personal communication). A further example of the convergence of lexico-semantic structure, found in Karaim but not in Tatar, is the extension of the use of

Karaim verbs to parallel the polysemy of equivalent Slavic verbs; for example Karaim *tur-* ‘stand’ has now been expanded in meaning to include ‘cost,’ just as the verb ‘to stand’ also means ‘to cost’ in some varieties of Slavic (Csató in press: 2).

In addition to the phenomena presented above, Karaim has been influenced by Slavic in several other ways that Tatar has not. The phonetic influence of Russian found in standard Tatar is still limited to the stratum of Russian borrowings – while earlier borrowings were adapted to Tatar phonotactics, later borrowings remain true to Russian pronunciation and phonological rules, even if they violate traditional Tatar phonology, e.g., have consonant clusters in syllable onsets.⁵ However Karaim, under the influence of Slavic, now has contrastive palatalized and unpalatalized consonants. Additionally, while Tatar has remained mostly verb-final, with adverbs and particles sometimes placed in the sentence-final slot for emphasis, Karaim word order is now “relatively free” with a tendency to SVO (as opposed to standard Turkic SOV), seen as an accommodation to the “European SOV-Sprachbund” (Csató in press: 3).

In conclusion, we have seen a variety of ways in which Karaim has been influenced by the Slavic languages with which it has been in contact. Some of these borrowing phenomena are parallel with those found in Tatar, while others indicate that Karaim is at a later, more intensely influenced stage of lexical and structural borrowing. It was seen that nine Slavic discourse-pragmatic words have become standard borrowings in Karaim, and that eight of these words (or their Russian equivalents, if they are Polish) are found code-

⁵ While the inability of some speakers to pronounce certain Tatar-specific phones is the subject of much commentary in the Tatar public ‘discourse of purity’, as discussed in Chapter Two, this Russian interference is still limited to a subset of speakers.

mixed in Tatar. While most of the non-content words borrowed from Slavic into Karaim are used in their discourse-pragmatic functions, the borrowing of non-discourse pragmatic prepositions (Polish *koło*, Russian *с* and *за*) indicates that Karaim is in the third stage of lexical borrowing in the model being laid out in this chapter. Other phenomena show a higher level of structural influence than associated with the second stage of lexical borrowing: the importation of Russian inflectional morphology; the use of Karaim inflectional morphology to fill grammatical requirements imported along with Slavic prepositions; the innovation of an instrumental case marker to parallel Slavic; a change to SVO word order; the innovation of the use of the enclitic *da* as a coordinating conjunction parallel to Slavic *и*; a new means of deriving compound nouns based on a Slavic model; and the innovation of the use of Karaim question words as relativizers, making them parallel in polysemy and function to their Slavic equivalents. None of these innovations are found in Tatar, which is only entering the second stage of borrowing from Russian – for example, Russian discourse-pragmatic words are still nonce borrowings in Tatar rather than standard borrowings. In the next section, an examination of Arabic and Persian borrowings in Tatar, we will see the vestiges of an earlier language contact situation where Tatar was influenced by its two literary superstrates.

4.3 Case Study Two: Arabic and Persian influences in Tatar

Up until the 19th century, the Tatars shared three literary languages with the Muslims of Russia: Classical Arabic, Classical Persian, and Chagatay (Rorlich 1986, Bennigsen and Quelquejay 1961). Arabic was a high language for Muslims of the Volga region for a millenium, from the 922 conversion of the Volga Bulgars to Islam to the anti-Islamic Soviet policies of the 1920s, and Classical Persian became a high language several centuries later. The mid-19th century saw the rise of a Tatar literary language based on the Kazan dialect; this language was used for education and literature, including newspapers and journals. The Tatar literary language was very heavily influenced by Arabic and Persian and was unintelligible to uneducated speakers of the vernacular – the relationship between 19th century literary and vernacular Tatar is in this way parallel to the relationship between 19th century Ottoman and vernacular Turkish. In fact, while modern post-lexical-reform Turkish and modern post-lexical-reform Tatar are not mutually comprehensible and must be learned as foreign languages, someone who knows Ottoman Turkish can read pre-Revolutionary literary Tatar almost without a dictionary, and vice versa. Literary Tatar does not appear to have been as heavily relexified with Arabic and Persian as Ottoman Turkish was, nor did it import grammatical rules such as gender agreement, but it was significantly influenced by these two superstrates. Lexical reform begun by Tatars in the late 19th century promoted the use of native Turkic words in literary Tatar, and the use of Arabic and Persian loanwords by certain authors decreased. Soviet-era language reform was aimed at removing Arabic and Persian loanwords and replacing them with their Russian equivalents. The end result is that Arabic and Persian influence in modern Tatar is significantly less than it was a century ago.

The analysis of Arabic and Persian influence on Tatar is necessarily different from the other case studies of lexical and structural borrowing presented in this dissertation, as here the languages in question are literary superstrates rather than dominant local vernaculars, and had a much greater effect on the written literary language than they did on the spoken language of the people. Thomason and Kaufman (1988: 78) noted that “minor structural influence from a prestigious literary language sometimes occurs through the written medium alone, without actual oral bilingualism among borrowing-language speakers.” Arabic is one of these influential languages, and has influenced the phonology and syntax of some Turkic languages in addition to the lexicon (*ibid.* 78-79). An examination of the influence of Arabic and Persian on Tatar shows that a language need not be contracting, nor in the process of multi-generational language shift, in order to borrow discourse-pragmatic words. Tatar speakers were not shifting to Arabic or Persian as their everyday languages; however, there was a distinct asymmetry in prestige and power, and Arabic and Persian dominated important functional domains, most notably education and literature, for many years. For post-imperial Turks, this dominance was enough for Arabic to be considered a threat, and post-Ottoman Turkish lexical reform was aimed at removing all Arabic and Persian influence from both written and spoken Turkish. As we will see, literary Tatar (but not vernacular Tatar) seems to have had Stage Three borrowing from Arabic and Persian, where grammatical words without discourse-pragmatic function were borrowed. However, as the intensity of contact lessened, so too did the influence of these two languages, and the only Arabic and Persian borrowings found in present-day Tatar, both written and spoken, are content words and discourse-pragmatic words (borrowings found in Stage One and Stage Two). This case study is divided into four major sections. In section 4.3.1, I present the Arabic discourse-pragmatic borrowings found in present-day Tatar, and

in section 4.3.2 I present the Persian discourse-pragmatic borrowings found in present-day Tatar. Section 4.3.3 is devoted to an analysis of Arabic and Persian discourse-pragmatic words in context, and uses excerpts from the post-Soviet Tatar press. Finally, section 4.3.4 examines the archaic Arabic and Persian borrowings that are no longer used in modern Tatar.

The etymologies of the Arabic and Persian loanwords cited here come from two major sources: Budagov's 1869 comparative dictionary of Turkic languages and Mäkhmütov et al.'s dictionary of Arabic borrowings, first printed ("with great difficulty") in 1965, and then reprinted in 1993. While some of the Arabic loanwords in question, particularly those limited to literary register, are perceptible to speakers of Tatar as loanwords, many of them are not perceived as such.

4.3.1 Arabic discourse-pragmatic words found in present-day Tatar

Although thousands of Arabic words and expressions were removed from Tatar during the Soviet era (Mäkhmütov 1993), many hundreds have remained. All of the 41 Arabic words or phrases in present-day Tatar that are not nouns, verbs, or adjectives (i.e., prototypical content words) are discourse-pragmatic words used in metalinguistic functions. Additionally, 27 of these words and phrases, which is to say, approximately two-thirds, have Russian equivalents that are code-mixed in Tatar-preferred style as described in Chapter Three. (For an at-a-glance summary of Arabic loanwords in Tatar and their code-mixed Russian equivalents, see Appendix C at the end of this dissertation.) In other words, many of the "native" Tatar equivalents of these Russian code-mixed words that are in the process of becoming standard borrowings are in fact themselves borrowings from Arabic – one wave

of discourse-pragmatics borrowing is supplementing, and perhaps replacing, an earlier one. Like the Russian code-mixed words in present-day Tatar, past Arabic borrowings fall into three major discourse-pragmatic categories: (1) Markers of discourse structure and force; (2) Interactional performatives, and (3) Evaluatives.

Markers of discourse structure and force

While there are no subordinating discourse markers from Arabic still used in present-day Tatar, there are coordinating and contrastive discourse markers. They are as follows:

[4.28]

- *wä* ‘and’
- *ämma* ‘but, however’
- *läkin* ‘but, however’
- *khal’buki* ‘whereas’
- *khayir* ‘however, although’

The coordinator *wä* ‘and’, which has a Russian equivalent that is quite commonly used in Tatar-preferred style (see section 3.4.1b of Chapter Three), is found in literary register only. Also found in literary register only are *khal’buki* ‘whereas’ and *khayir* ‘however’. *Khayir* was originally borrowed with a meaning of ‘no’ (Mäkhmütov et al. 1993), which is its meaning in present-day Turkish, and seems to have undergone a semantic shift and rather than expressing ordinary negation it is used only to mean ‘however’ and ‘although.’ The two other contrastive discourse markers, *ämma* and *läkin*, both meaning ‘however’ (and sometimes used in combination, *ämma läkin*, also meaning ‘however’), are quite commonly found in all registers of Tatar, and their Russian equivalent *no* ‘however’ is quite commonly code-mixed in Tatar-preferred style (see section 3.4.1c of Chapter Three).

Eight Arabic discourse-pragmatic words are used as metacommentary in present-day

Tatar:

[4.29]

- *älbättä* ‘of course’
- *fälän* ‘such and such’
- *jämägat* ‘people, folks’
- *makhsus* ‘especially, particularly’
- *mäglüm (ki)* ‘it is well known (that)’
- *mäsälän* ‘for example’
- *nihayat* ‘finally!’
- *wässäläm* ‘and that’s it!’
- *yag”ni* ‘that is, i.e.’

Three of these words are used in literary register only: *fälän* ‘such and such’, *mäglüm (ki)* ‘it is well known (that)’ (where the *ki* is a borrowing from Persian), and *wässäläm* ‘and that’s it!’.

Of these three words, only *wässäläm* has a Russian code-mixed equivalent, *i vse!* – this Russian phrase is not in any way marked for register (for more on *i vse!* and other Russian code-mixed metacommentary words and phrases, see section 3.4.1d of Chapter Three). Arabic *jämägat* ‘people, folks’ is used as a directive in both speech and writing to both address the audience and focus attention on the following proposition, what Fraser (1996) calls a vocative marker. It is slightly more formal than Russian *rebiata* (originally the just the plural of *rebionok* ‘child’, now used to colloquially address a group of young peers) and slightly less formal than English *ladies and gentlemen*. The most textually oriented words, *mäsälän* ‘for example’ and *yag”ni* ‘that is’, are found more frequently in writing, but are not limited to literary register, and are quite commonly used. Their code-mixed Russian equivalents, *naprimer* ‘for example’ and *to est* ‘that is,’ are also common.

Interactional performatives

Tatar has eight words and phrases borrowed from Arabic that function as interactional performatives:

[4.30]

- *sälam* ‘hi, hello’
- *märkhaba* ‘hello, welcome’
- *ässälame galäykem* ‘hello’
- *räkhim it(egez)* ‘welcome’
- *khush* ‘bye’
- *räkhmät* ‘thank you’
- *gafu it(egez)* ‘pardon me, excuse me’
- *mökhtäräm* ‘respected’

Three of the four Russian interactional performatives code-mixed in Tatar (see section 3.4.2 of Chapter Three) are use in lieu of Arabic loanwords rather than native Tatar terms. Russian *privyet* ‘hi’ replaces Arabic *sälam*, Russian *poka* ‘bye’ replaces Arabic *khush* (among other leave-takings, some of which are entirely Tatar), and Russian *izvini(te)* ‘excuse me’ replaces *gafu it(egez)*, where the Arabic noun *gafu* ‘pardon’ is transformed into a verb with the Tatar “nativizing” verb *iterge* ‘to do, to make’ (such that the literal translation of *gafu it(egez)* would be something like ‘do a pardon’). These three Arabic borrowings are among the most commonly used forms of greeting, leave-taking, and asking pardon (and *gafu it(egez)* is really the only way that one excuses oneself in Tatar).

Giving thanks in Tatar is also done with an Arabic loanword, *räkhmat*, which comes from the same Arabic root as the word *rakhim*, nativized with the verb *iterge* ‘to do’ to form the phrase *rakhim it(egez)* ‘welcome,’ used to greet guests. (The other standard expression used to greet guests uses Arabic *khush*: *khush kildegez*, which translates literally as ‘you came well’.)

The remaining interactional performatives are less commonly used, and marked either socially or in register. The greeting *ässälame galäykem*, formulaic in Arabic, is sociolinguistically marked and used only by religiously observant or politically oriented/nationalistic Tatars, while *märkhaba* ‘hello, welcome’, which is a standard greeting in Turkish, is limited to literary register only in Tatar and is very infrequently used. Finally, *mökhtäräm* ‘respected’ is used in combination with a first name, and sometimes also patronymic, to form a polite address term, e.g., *mökhtäräm Siuzanna*⁶ – it can thus be seen as an expression of attitudinal deixis, “the use of indexical expressions which signal aspects of social status and/or forms of respect” (Verschueren 1999: 20-21). *Mökhtäräm* is rather infrequently used and marked as quite formal, and thus very polite. Far more common as a polite form of address is the combination of first name followed by words used to denote respect: *khanym* for married women, *tutash* for unmarried women, *apa* (literally ‘aunt’ or ‘older sister’) for women older than oneself with whom one is on familiar terms; *äfändi* for most men, and *abiy* or *abziy* (literally ‘uncle’ or ‘older brother’) for men older than oneself with whom one is on familiar terms.

⁶ This is the way one young Tatar begins his e-mails to me.

Evaluatives

Discourse-pragmatic Arabic loanwords are used in present-day Tatar for positive evaluation, as epistemics, and to express evaluations of time, degree, and relationship to reality. The three positive evaluations are all interjections, and all are quite common:

[4.31]

- *afärin* ‘bravo! well done!’
- *amin!* ‘amen’
- *insha alla, inshalla* ‘God willing’

The Russian equivalent of *afärin* ‘bravo! well-done!’ is *molodets*, frequently code-mixed in both speech and informal writing. Arabic *amin* is less-frequently used, although it is not limited to strictly religious contexts. (And it should be noted that *amin* or its Hebrew cognate is a discourse-pragmatic term of positive evaluation that has been borrowed into a multitude of languages, including English, Russian and Turkish, to name just three.) Unlike *amin*, *insha alla*, or as is more common, *inshalla* ‘God willing’, is religiously marked, and used by more religiously observant or politically oriented Tatars.

The most frequently used epistemics in present-day Tatar are in fact Arabic loanwords:

[4.32]

- *bälki* ‘probably; perhaps, maybe’
- *ikhtimal* ‘probably’
- *mögayin* ‘probably’
- *mömkin* ‘possible, probable, likely’

The Russian equivalents of these loanwords are frequently code-mixed in Tatar-preferred style: Russian *naverno* ‘probably’ is used in lieu of the very common Arabic *bälki* (used in all registers), *ikhtimal* (more literary), and *mögayin* (more literary). In fact, of the four words used

to mean ‘probably’ in present-day Tatar, only *sbayat*’ is not an Arabic loanword. Additionally, when Russian *mozhet* ‘might’ is code-mixed, it is used in lieu of *mömkin*, itself a borrowing from Arabic (for more on the code-mixing of Russian epistemics, see section 3.4.3d of Chapter Three).

The largest category of borrowed Arabic discourse-pragmatic words in Tatar, as with code-mixed Russian words (section 3.4.3e of Chapter Three), is that of evaluations of time, degree, and relation to reality. There are two loanwords for evaluations of time:

[4.33]

- *daimi* ‘constantly, continually’
- *gomumän* ‘usually, in general’

These two words are used in all registers of Tatar. While I have yet to find an instance of code-mixed Russian *postoyanno* ‘constantly, continually’, which would be used in the place of *daimi*, *voobshche*, which is the Russian equivalent of *gomumän* ‘usually, in general’, is very common in Tatar-preferred style.

Many, although not all, of the Arabic loanwords used for evaluations of degree are intensifiers:

[4.34]

- *fäkat*’ ‘only, just’
- *föj’ätän* ‘unexpectedly’
- *gajäp* ‘surprisingly’
- *gayat* ‘very’
- *ifrat* ‘overly’
- *kadär* ‘as much as’
- *khätta* ‘even’
- *khosusän* ‘especially, particularly’
- *tämamän* ‘entirely, completely’

Four of these nine words have code-mixed Russian equivalents: *prsto* ‘only, just’ (used to express limitation, not as a hedge) is used in lieu of *fäkat*’; *slishkom* ‘overly’ is used in lieu of

ifrat, *daşbe* ‘even’ is used in lieu of *kbätta*; and *sovsem* ‘completely’ is used in lieu of *tämamän*. The adverbs *föj’ätän* ‘unexpectedly’ and *kbosusän* ‘especially, particularly’ are found in literary register only, and *föj’ätän* is so infrequently used that it can now almost be classified as archaic. The postposition *kadär* is currently used in Tatar most commonly in its spatial sense, meaning ‘up to, or until’, but it was most likely first borrowed in its evaluative meaning of ‘as much as’, just as Russian *okolo* ‘around’ is currently code-mixed only in its evaluative form as a hedge and is not yet used as a spatial relator meaning literally ‘around’. The fact that Arabic *kadär* is used with both its root meaning and grammaticalized meaning is indicative of the length and intensity of Arabic influence on written Tatar, and suggests that other grammatical words without discourse-pragmatic functionality were borrowed into Tatar from Arabic. This is indeed the case, although none of these grammatical words are used in present-day Tatar – they will be presented and discussed in Section 4.3.4 below, which is devoted to now-archaic Arabic loanwords.

Finally, there is one Arabic loanword used to evaluate the relationship to reality:

[4.35]

- *motlakan* ‘absolutely’

This adverb is rarely used, and is limited to the literary register. However, its Russian equivalent, *absolütno*, is frequently code-mixed in Tatar-preferred style.

A final note of interest – many of the Arabic discourse-pragmatic borrowings listed above are also found in Persian, which borrowed extensively from Arabic, and include: *va* ‘and’ (Tatar *wä*), *ammaa* ‘as, for, but’ (Tatar *ämma*), *laakin* ‘but’ (Tatar *läkin*), *hattaa* ‘even’ (Tatar *kbätta*), *fagat* ‘only’ (Tatar *fäkat*), and *daa’iman* ‘continually’ (equivalent to Tatar *daimi*) (Versteegh 1997).

4.3.2 Persian discourse-pragmatic words found in present-day Tatar

Persian borrowings are often considered to be part of the stratum of Arabic borrowings in Tatar, even by Tatar lexicographers. For example, the *Garäpçä-Tatarçha-Ruscha Alınmalar Sözlöge* ‘Arabic-Tatar-Russian Dictionary of Borrowings’ that is the source for the majority of the etymologies presented here includes borrowings from Persian, even though they are not mentioned in the title of the dictionary – loanwords of Persian origin are simply marked with a barely-noticeable “f” for *farsi* after the headword of the entry. Persian loanwords make up between ten and twelve percent of this stratum of borrowings (Mäkhmütov 1993: 798), and there are correspondingly fewer Persian discourse-pragmatic loanwords used in present-day Tatar: 13, as opposed to 41 from Arabic. Unlike many of the Arabic borrowings discussed above, none of the Persian discourse-pragmatic loanwords found in present-day Tatar are restricted to literary register or sociolinguistically marked in any way. Additionally, their status as borrowings is generally not known by the average Tatar speaker – they are used as fully native words. Finally, some of these borrowings, in particular several coordinators, have expanded beyond merely metalinguistic function and are now also used in their root meanings or purely grammatical functions, as will be seen below. Seven of these thirteen loanwords, or 54%, have Russian equivalents that are code-mixed in Tatar- preferred style. A comparison of the use of Persian loanwords and Russian code-mixed words in Tatar gives further evidence for the hierarchical nature of the model of lexical and structural borrowing and presented here: when they are first being borrowed, discourse-pragmatic words will be used only in their metalinguistic and evaluative meanings, and only after time will they begin to be used in their root meanings and grammatical functions.

The majority of these Persian discourse-pragmatic borrowings are used to structure discourse. There are three subordinating discourse markers:

[4.36]

- *ägär* ‘if’
- *chönki* ‘because’
- *ki* subordinate clause indicator, can often be glossed in English as ‘that’

The borrowing Persian *chönki* ‘because’ is associated with syntactic alteration: Tatar sentences with Persian *chönki* do not have standard Turkic-style pre-head embedded subordinate clauses, but rather Indo-European style post-main clause subordinate clauses. While *chönki* has been described as used mostly in Tatar’s literary register (R. R. Shamsutdinova, personal communication), it is actually quite common in present-day youth speech – perhaps because of its syntactic equivalence with Russian *potomu chto*. The subordinating *ki*, used in writing only, is interesting not only because its use too introduced post-main clause subordination, but also because it is apparently in the process of being reinterpreted as an emphatic particle; it was defined for me by several young Tatars as giving emphasis to or heightening what was being presented, an interesting semantic reinterpretation where a metalinguistic word originally used to structure discourse becomes a metalinguistic word used for evaluation. An example of *ki* used in a context with an ambiguous reading, where this reinterpretation is clearly possible, can be found in section 4.3.3 below.

The final subordinator, *ägär* ‘if’, can also be seen as both structural and emphatic. The use of *ägär* is optional, as the most common native way to express conditionality in Tatar is with the verbal conditional affix *-sA-*, e.g., *bar-sa-m* ‘if I go,’ (where *bar-* is ‘go’, *-sa-* is the conditional, and *-m* is the first person singular ending) and *kil-sä-ng* ‘if you come’ (where *kil-* is ‘come,’ *-sä-* is the conditional, and *-ng* is the second person singular ending). A conditional clause using *ägär* is double-marked for conditionality: *ägär* will be the first word

of the clause (usually the first word of the sentence), while the verb in the conditional will be the last word of the clause, as Tatar is a verb-final language. The use of *ägär* can be seen as both grammatical, in that it denotes conditionality, and as emphatic – not only is the conditionality of a proposition marked twice, but it is now marked both at the first and last words of a clause, traditional places of emphasis.

Tatar borrowed a single coordinative conjunction from Persian:

[4.37]

- *häm* ‘and’

The borrowing of this conjunction was syntax-altering. Tatar’s native Turkic coordinators are not conjunctions, but rather the enclitic *DA* – which assimilates in voicing and vowel harmony to the preceding word – and the postposition *belän* ‘with’. *DA* can be used to coordinate both nouns and verb phrases, while *belän* is used only used in the coordination of nouns. Tatar sentences using *häm* as a coordinator have a different syntactic realization from those using *DA* and *belän*. (1) *Häm* is used once in a sentence to coordinate nouns or propositions, while *da* is used twice, once after each of the nouns or verb phrases being coordinated. (2) While the construction ‘X *häm* Y’ appears at first glance to have the same syntactic structure as ‘X *belän* Y,’ it does not: ‘X *häm* Y’ is best translated as ‘X and Y,’ while ‘X *belän* Y’ is actually ‘Y and X,’ where X is the more adjunct item, e.g. *limon belän chey* ‘tea with lemon’ (lit. ‘lemon with tea’). The coordination in a Tatar phrase or sentence using the borrowed conjunctions *wä* or *häm* is syntactically identical to the coordination in a phrase or sentence using the code-mixed Russian equivalent, *i*. *Häm* is used in present-day Tatar not just metalinguistically, to coordinate propositions or “idea units” (Schiffrin 1987), but also in its root grammatical function, to coordinate individual items. This can be seen in context in

section 4.3.3 below. Its code-mixed equivalent, Russian *i*, is used only as a discourse-pragmatic marker coordinating idea units.

There are five contrastive discourse markers borrowed from Persian that are still used in present-day Tatar:

[4.38]

- *gärchä* ‘although’
- *mägär* ‘however,’ ‘only’
- *ya* ‘or’
- *yakbut* ‘or’
- *yaki* ‘or’

The use of *gärchä* ‘although’ is much like that of *ägär* ‘if’ as described above. As with *ägär*, the use of *gärchä* is optional – Tatar has a native morphosyntactic means of creating concessive clauses through the combination of a verb in the conditional (with the affix –*ɫ1*-) and the enclitic *da*, found clause finally, for example:

[4.40] The Tatar concessive construction

Aprel’ urta-lar-ı bul-sa da, kön suık.
April mid-PL-POSS be-COND ENC day cold
‘Although it is mid-April, the day is cold.’ (Kefeli 1999: 211)

If placed at the beginning of this clause (*Gärchä* Aprel’ urtalari bulsa da...), *gärchä* would alter the structure of the clause by introducing the element of concession at the beginning, and also would serve to emphasize the proposition by means of double-marking the concession.

The concessive *gärchä* is not used nearly as frequently as conditional *ägär*, and also infrequently used is Persian *mägär* ‘but, however,’ although its Russian equivalent, *no* ‘but, however’ is frequently code-mixed in Tatar. The three Persian conjunctions meaning ‘or’, *ya*, *yakbut*, and *yaki* contrast with their code-mixed Russian counterpart, *ili*, in two main ways.

The first is that the Persian borrowings are used more frequently in literary registers, while Russian *ili* is only in casual registers. The second is that while Russian *ili* is only used metalinguistically, to contrast propositions, the Persian conjunctions are used to contrast both propositions and individual items – this can be seen for Persian *yaki* in context, shown below in section 4.3.3, and is parallel to the distribution of *häm* ‘and’, which is also used both discourse-pragmatically and grammatically.

The final Persian discourse-pragmatic words still used in present-day Tatar are evaluatives. There are actually very few Persian evaluatives that are still used: there are no Persian borrowings used for positive evaluation, for negative evaluation, as epistemics, or as hedges. There are, however, three quantifiers used to evaluate degree:

[4.41]

- *hämmä* ‘all, each’
- *här* ‘each, every’
- *hich* ‘nothing, none’

Most interesting of these three words is Persian *här*, as it has become part of standard compound words in Tatar, for example:

[4.42]

- *härber* ‘each’ (*ber* = Tatar ‘1’)
- *härchakta* ‘at any time, always’ (*chakta* = Tatar ‘time’ in the locative case)
- *härkayda* ‘everywhere’ (*kayda* = Tatar ‘where’)
- *härkaysi* ‘each of them’ (*kaysi* = Tatar ‘which of them’)
- *härkaychan* ‘always’ (*kaychan* = Tatar ‘when’)
- *härkem* ‘each[person]’ (*kem* = Tatar ‘who’)
- *härkhäldä* ‘in any case’ (*khäldä* = Tatar ‘state’ in the locative case)
- *härtörle* ‘each, each kind’ (*törle* = Tatar ‘various’)
- *härwakit* ‘always’ (*wakit* < Arabic, = Tatar ‘time’)

Note that all of these compounds are themselves metalinguistic in function, and the majority of the Tatar words used in conjunction with *här* to form compounds are themselves function words.

4.3.3 Arabic and Persian discourse-pragmatic words in context

In this section, I will provide three examples of Arabic and Persian discourse-pragmatic words used in context. The examples are three brief paragraphs, each several sentences long, that are stretches of discourse typical of the three articles from which they were extracted, articles with various levels of scholarliness and formality. Although many of the Arabic discourse-pragmatic loanwords described in the previous sections are limited to the literary register, it will be seen that the frequency of metalinguistic Arabic and Persian borrowings does not necessarily decrease because the discourse in question is less literary or less formal in tone. The excerpts are also interesting because they demonstrate the extent to which the present-day Tatar lexicon is composed of borrowings, some of which are perceptible to speakers as loanwords, and some of which are not.

The first excerpt comes from a scholarly article on the return of Arabic borrowings to post-perestroika Tatar (it is interesting to note that in the article, the loanwords discussed are almost exclusively nouns, even when the discussion turns to well-established Arabic loanwords that are generally felt to be native). The article was published in the general scholarly journal *Fän häm Tel* ‘Science and Language’, which is devoted to the humanities, social sciences, and hard sciences, and it is written in *fänni stile*, academic style. In the Tatar text, borrowed discourse-pragmatic words are in **boldface**, borrowed Arabic content words are in *italics*, and native discourse-pragmatic words are underlined. (The same is true of their glosses in the English translation.)

“**Ämma** teldä tege yaki bu töshenchäne yish kına barı tik alınma süz genä belderergä **mömkın**. **Mäsälän**, tatar telendä *kitap, däftär, mäktäp, tärtip, tarıkb* kebek töshenchälärne belderä torgan süzlär yuk. Alar barı tık shushı alınma (garäptän kergän) süzlär belän genä belderelälär... Bez yish kına bu süzlärneng alınma buluın da sızımbez...” (Safiullina and Fyodorova 2000: 17).

“**However**, it is often possible that in a language this or that concept can only be expressed with a loanword. **For example**, in the Tatar language there are no words to express concepts such as *book, notebook, school, upbringing, and history*. They can only be expressed with these loanwords (which have come from Arabic)... And we quite often do not feel that these words are borrowings...”

There are four borrowed discourse-pragmatic words in this paragraph: 1. **ämma** ‘however’, 2. **yaki** ‘or’ (the only discourse-pragmatic word here from Persian), 3. **mömkın** ‘possibly, maybe’, and 4. **mäsälän** ‘for example’. The only other Arabic borrowings in the excerpt are those quoted as “native-seeming” loanwords. Additionally, there are eight other discourse-pragmatic words found in this excerpt: 1. tege ‘that’, 2. bu ‘this’, 3. yish ‘often’, 4. kına which is used both as ‘only’ and as an emphatic particle (due to consonant and vowel assimilation, it is also realized in this excerpt as genä), 5. yuk ‘there is/are not’, 6. kebek ‘such as’, 7. tık ‘just, only’ and 8. da ‘and’. So in total we find twelve discourse-pragmatic words in this excerpt, three of which are from Arabic, one which is from Persian, and eight that are native. Note that the Arabic loanwords, **ämma**, **mömkın**, and **mäsälän**, are all in peripheral locations: they are the first and last words of the first sentence, and the first word of the second sentence. The Persian coordinator **yaki** is more integrated in the sentence, since it is being used here as a general coordinating conjunction. Of the Tatar discourse-pragmatic words found in the excerpt, only one is located peripherally: yuk ‘there are no’, which serves as the copula and thus is sentence-final in accordance with standard Tatar word order, which is verb-final. Additionally, the Arabic discourse-pragmatic words appear to be peripheral to consciousness as well: while the article is devoted entirely to the use, past and present, of

Arabic loanwords, none of the 41 Arabic discourse-pragmatic words used in present-day Tatar (eleven of which are used in the article itself) are mentioned or discussed in any way.

The second excerpt is taken from an article printed in the daily government newspaper *Watanım Tatarstan* ‘My Homeland Tatarstan’, which is the continuation of one of Tatarstan’s two Soviet-era newspapers and has a wide general readership. The article is therefore aimed at a general rather than a scholarly audience and is thus written in a less formal and scholarly style; for example, the author uses the first-person singular to refer to himself, where a more scholarly style would require use of the first-person plural. Even so, the article is somewhat pedantic and lecturing in tone, and devoted to summarizing and critiquing problems with modern Tatar usage as found in the press and on television. In addition to the presentation conventions used above, small capitals are used for Russian loanwords and their glosses, and native inflectional and derivational morphology suffixed to loanwords is left in plaintext both in Tatar and in the translation:

“...khäzer **här** GAZETA, ZHURNAL üzeneng REDKOLLEGIYÄse SOSTAVına **bik** kup *därjäle shäkebeslärne* kertä bashladı. Kayberlärendä **khätta** ike MINISTR, bernichä AKADEMIK, *fän* DOKTORLARİNİNG *isem*-FAMILİYÄläre kiterelä. **Ägar** alar REDAKTSIYA eshendä nindi dä bulsa kingesh belän, *tänküydi fiker* belän katnashsalar, ber **genä** süzem dä **yuk**, moni yakhshı kürenesh, dip saniym... **Ägär** **dä** **mägär** ukıysalar, **här** bitendä yalgışlar tulıp yatkan GAZETA-ZHURNALNING *därjäle* “tuy GENERALI” buludan oyalırlar ide... **Chönki** *mäg* “näsezleklär, yalgışlar, ber süz belän äytkändä, telebezne bozu öchen, REDKOLLEGIYÄ *äg*” *çalari* bularak, alar **da** *jawaplıdır* BIT **inde!**” (Safiullin 1999, ellipses original).

“...now **each** NEWSPAPER and JOURNAL has started to bring **very** many *authority figures* into their EDITORIAL STAFFS. In several of them are found **even** the *NAMES* of two MINISTERS, several ACADEMICS, and *Ph. D.s*. **If** they are participating in EDITORIAL work with some kind of advice or *critical thought*, then I have nothing to say (lit. **there is not even** one word of mine), I consider this a good thing... **But if** (lit. **if and but**) they read how **each** page is filled with mistakes, they would be embarrassed to be authoritative ‘wedding GENERALS’ of these newspapers and journals... Because as *members* of EDITORIAL STAFFS, they **too** are *answerable* after all for the *meaningless* items and mistakes, in a word, for the ruining of our language!”

All but one of the Russian borrowings in this excerpt are nouns that relate to the media, education, and politics – semantic fields that were heavily relexified with Russian words during the Soviet era (and not coincidentally, fields that were dominated by Russians). They are as follows: GAZETA ‘newspaper’, ZHURNAL ‘journal’, REDKOLLEGIYÄ ‘editorial staff’, SOSTAV ‘staff’, MINISTR ‘minister’, AKADEMIK ‘academic’, DOKTOR ‘Ph.D.’ (in combination with *fän*), FAMILIYÄ ‘last name’, and GENERAL ‘general’. Note that many of these words are themselves borrowings from other languages, often referred to in Tatar as “internationalisms” – however, their etymologies are usually not known by the average Tatar speaker; they are felt to be Russian.

The majority of the Arabic borrowings found in the excerpt are also nouns (although several are suffixed with native Tatar derivational morphology) and also belong to semantic fields typically associated with superstrate languages: *därjä* ‘authority’, *shäkebes* ‘person, individual’, *fän* ‘science’ (here combining with Russian DOKTOR to mean Ph.D.), *isem* ‘first name’, *tänkäydi* ‘critical’, *fiker* ‘thought’, *mäg’nä* ‘meaning’, *äg’za* ‘member’ and *jawab* ‘answer’. Particularly interesting are *äg’za* ‘member’, which is a word very consciously used in post-Soviet Tatar in lieu of the Russian borrowing *chlyen*, and *tänkäydi* ‘critical’, which uses Arabic derivational morphology to form the adjective rather than Tatar – this form is not found in modern Tatar dictionaries, and in standard Tatar the concept of ‘critical’ is expressed not with an adjective, but with a nominal compound composed of the borrowed Arabic noun *tänkü* ‘criticism’ and the noun it is modifying marked with the *izafet* suffix of *-(s)I*. These two words mark the author (or perhaps his editor, as seen in Chapter Two) as a person who is conscious of the Arabic provenance of the words he is using, and in favor of their usage. Even so, this does not lead to a higher level of usage of Arabic discourse-pragmatic words

than found elsewhere: there are five borrowed discourse-pragmatic words in the excerpt, and only one of them as Arabic – the other four are Persian. Arabic **khätta** ‘even’ and Persian **mägär** ‘but’ and **chönki** ‘because’ are used one time apiece, while Persian **här** ‘each’ and **ägär** ‘if’ are used twice apiece. Several of the five native discourse-pragmatic words found here are also found in the first excerpt: genä as an emphatic particle, da to mean ‘and’, and yuk ‘there is not’. The other two native discourse pragmatic words are the intensifier bik ‘very’ and the emphatic particle inde, which is not glossable in this text. Finally, the emphatic particle BIT, used in this excerpt in combination with inde, is the borrowed form of the Russian particle *ved’* – while *ä*, the nativized form of Russian *a* ‘and/but’ is not present in this excerpt, it is used to introduce several sentences elsewhere in the article (for more on these two borrowings see section 3.3 of Chapter Three). To summarize in numerical terms, in this excerpt there are 82 words, or 71 tokens, once repeated words are removed. Of these tokens, 60 are denotative content words, and 11 are metalinguistic discourse-pragmatic words. Of the content words, nine are Arabic loanwords (15%), ten are Russian loanwords (17%) and 41 are native stock (68%). Of the discourse-pragmatic words, one is Arabic (9%), four are Persian (36%), one is Russian (9%), and five are native (46%). The discourse-pragmatic words of foreign origin are once again peripheral with regard to word order, and in the two cases in the excerpt when they are not the first word of a sentence or clause, they are preceded only by a single contextualizing word: *khazär* ‘now’, and *kayberlärendä* ‘in some of them’.

The third and final excerpt comes from an article that is not at all scholarly and was printed in the weekly newspaper *Tatar Ile* ‘Tatar Country’, which has a wide readership both within and outside of Tatarstan. This article too is written in first-person singular, and is less

formal in tone (although not informal), as befits the subject, which is in no way academic but rather the recounting of the unexpected visit of a foreigner at the newspaper's offices.

“Berazdan bez aning belän shulkadär yakınlaşip kittek **ki**, ul **khätta** REDAKTSIYÄbez urnashkan bülmägä kerergä *riẓa*lıgın birde. Biredä, yegetlär **häm** kızlar belän utırıp alingan kilesh, **härkemneng** sorawına tögäl **häm** achik *jawap* birergä tırıştı” (Sälimov 2001)

“In a little while, we had become **so** well-acquainted (lit. had become close **as much as** that) **that** she **even** agreed to come into one of the rooms of the EDITORIAL OFFICE. There, while sitting with the young men **and** women, she tried to give a precise **and** open *answer* to the questions of **each** one.”

While the number of content-word borrowings found in this excerpt is quite low (two from Arabic, one from Russian), the percentage of borrowed discourse-pragmatic words is quite high. The two Arabic words, *riẓa* ‘agreement’ and *jawap* ‘answer’, are standard borrowings, unmarked for register or style, and the Russian word REDAKTSIYÄ is also a standard borrowing (used twice in the second excerpt above) for which there is no native equivalent. There is just one native discourse-pragmatic word used in the paragraph, berazadan ‘in a little while’, used to create temporal narrative structure, but there are five borrowed discourse-pragmatic words, two of which are used as one part of a compound word. Arabic **kadär** ‘as much as, up to’ is combined with native Tatar *shul* to give a literal meaning of ‘as much as that’, usually better glossed as ‘so’. This combination of Arabic *kadär* with a word of native stock meaning ‘that’ to mean ‘as much as that’ is not limited to Tatar – the word *oadar* is found in Asia Minor Greek, to be discussed below, and is clearly a phonetically reduced version of combined Turkish *o* ‘that’ and Arabic *kadar* (Dawkins 1916). Persian **ki** is used as a subordinate clause indicator, in this instance glossable as ‘that’, but this particular context is compatible with its ongoing reinterpretation as an emphatic particle rather than subordinating coordinator. Its syntactic position in this sentence is identical to how Russian

çhto ‘that’ would be used had it been code-mixed in this sentence, except that punctuation conventions are different – *ki* precedes the comma and *çhto* follows it. Arabic **khätta** ‘even’ is used quite standardly, as is Persian **häm** ‘and’, here used not to link propositions, but merely to coordinate two nouns (*yegetlär* ‘boys’ and *kižlar* ‘girls’) and two adjectives modifying the same noun (*tögäl* ‘precise’ and *achik* ‘open’). Finally, Persian **här** ‘each’ is found in the standard compound *härkem*, combining with Tatar *kem* ‘who’ to mean ‘each [person]’.

Several conclusions can be drawn from these three excerpts. The first is that a change in the level of scholarliness or in subject matter does not seem to change the number of discourse-structuring Arabic and Persian loanwords. Secondly, while native discourse-pragmatic words are also used for metalinguistic function, there are stretches of text when only Arabic or Persian discourse-pragmatic words are used. Thirdly, Arabic and Persian discourse-pragmatic words tend to be in the peripheries of sentences and clauses, and are most often sentence- or clause-initial, while Arabic and Persian borrowings that are now used as grammatical words without discourse-pragmatic functionality are found clause-internally. (Recall that, as described in Chapter Three, peripherality is traditionally seen as a typical characteristic of discourse markers – but note that the peripheral discourse-pragmatic words in these excerpts include, but are not limited to, discourse markers.) Finally, these excerpts show to what extent Tatar discourse is composed of borrowings, both content and discourse-pragmatic – even in written texts that have been edited in an era of lexical reform that has an explicit ideal of “pure” Tatar, it is difficult to find a sentence without a borrowing of some kind.

4.3.4 Archaic Arabic and Persian borrowings

Pre-Revolutionary literary Tatar, particularly before the de-Arabicization lexical reforms undertaken by Tatar writers themselves, was very heavily influenced by Arabic and Persian: for example, in pre-Revolutionary Tatar poetry, Persian and Arabic borrowings could make up as much as 65 percent of the lexical content (Mäkhmutov 1993). In this section, I present the Arabic and Persian discourse-pragmatic loanwords that are now archaic, whether due to conscious lexical reform or due to less-conscious processes of obsolescence. Because these words have been culled from dictionaries only, and not from texts, it is not clear how frequently or in what contexts they were used. However, for the purposes of this model of lexical borrowing, far more important than the distribution or diffusion of a loanword is the fact that it was borrowed at all. Even if some of the Arabic and Persian words found in the Tatar dictionaries used for this section were given entries although they were found in just one pre-Revolutionary Tatar text, they can then be interpreted as code-mixed, much like the Russian discourse-pragmatic words found code-mixed in the e-mails and website postings of young Tatars. In the model of lexical and structural borrowing and language presented here, code-mixing is the hypothesized mechanism by which discourse-pragmatic words become standard borrowings; therefore, the lack of information on just how standard or frequent these borrowings were does not diminish their usefulness as additional evidence for the model.

Some of the Arabic and Persian borrowings presented here have entries (where they are marked as archaic) in Soviet-era and post-Soviet era Tatar dictionaries (e.g., *Tatarsko-Russkii Slovar'* 1966), but the vast majority can only be found in dictionaries specifically devoted to Arabic and Persian borrowings. There are 142 archaic discourse-pragmatic words

from Arabic and 28 from Persian. When there is a code-mixed Russian equivalent for an archaic Arabic or Persian loanword, it is noted at the end of each subsection. Finally, at the end of this section, the 23 archaic grammatical words borrowed from Arabic and from Persian will be presented and discussed.

4.3.4a Archaic Arabic borrowings

The subordinating conjunctions borrowed from Arabic were:

[4.43]

- *binabärin* ‘therefore, in view of this’
- *fälihaça* ‘and therefore’
- *hakäça* ‘thus, in that way’
- *käça* ‘thus, in that way’
- *käçalik* ‘thus, in that way’
- *lihaçä* ‘therefore, on this basis’
- *liçalik* ‘therefore, for this’
- *räsmä* ‘thus, in that way’
- *täki* ‘in order to’
- *zirä* ‘because, since’

The Russian equivalents of *täki* ‘in order to’ (*chtoby*) and *zirä* ‘because’ (*potomu chto*) are both code-mixed in Tatar-preferred style.

Coordinating discourse markers were:

[4.44]

- *äyçan* ‘also, and’
- *niz* ‘also, and’

Russian *toçbe* is the equivalent of both of these terms, and code-mixed in Tatar.

Contrastive discourse markers were:

[4.45]

- *äm* ‘or’
- *ilłyä* ‘but, however’
- *wäläkin* ‘but, however’
- *wäli* ‘but, however’
- *hya äkalle* ‘although’
- *wäläü* ‘although, and though’
- *räg’män* ‘despite’
- *wä ilłyä* ‘otherwise’

Russian *no* ‘however’ is the equivalent of *ilłyä*, *wäläkin*, and *wäli*, and is found code-mixed in Tatar. The dictionary entry for *äm* ‘or’ states that it was used for “alternative propositions” (Mäkhmütov et al. 1993: 736); therefore it would seem that this word – like its Russian counterpart *ili* when code-mixed in Tatar – was used only in a discourse-pragmatic function.

The largest number of discourse-structuring words were used as metacommentary:

[4.46]

- *kyäshki* ‘if only, would that’

- *wä gaire zalik* ‘and so on, etcetera’
- *ila akbiribi* ‘and so on, etcetera’
- *ilkh* ‘and so on, etcetera’
- *wä bakäza* ‘and so on, etcetera’
- *hälömmä järä* ‘and so on, etcetera’

- *wäjbän* ‘on the one hand’
- *mäga ma fib* ‘together with this’
- *mäga haza* ‘together with that’
- *mägan* ‘together with...’
- *magada* ‘except’
- *ma siva* ‘except’

- *finäfsihi* ‘basically’
- *mälän* ‘not literally’
- *kbölasan kälyäm* ‘in brief’
- *älkbasıyl’* ‘in brief’
- *kbalänki* ‘the fact of the matter is...’
- *ätäm* ‘more precisely’
- *gayanän* ‘clear, clearer than clear’

- *biäyikhalèn* ‘in any case, at any rate’
- *galä köllettäk”dir* ‘in any case, at any rate’
- *ikhtiyatan* ‘anyway’

Turkish *kesbke* is the equivalent of the now-archaic Tatar *kyäshki*, and is still used in present-day Turkish to introduce wishes (and not merely in literary register). *Koroche*, which is the Russian equivalent of *khölasan kälyäm* and *älkhasiy!* ‘in brief’, is frequently code-mixed in Tatar-preferred style.

A great number of the discourse-pragmatic words used for metacommentary were deictic words:

[4.47]

- *hazä* ‘this’
- *äl’an* ‘now’
- *älkbal’* ‘now’
- *khaläya* ‘now’
- *äl’yaum* ‘today, now’
- *lyakhikan* ‘now’
- *ämma bäg”dä* ‘later, afterwards’
- *bäg”dä* ‘after’
- *möäkhkhärän* ‘later, afterwards’
- *wä bägüdä* ‘after that, later’
- *kablä* ‘earlier’
- *mökaddämä* ‘before, formerly’
- *därkbal’* ‘right away, quickly, immediately’
- *äüwälän* ‘first of all, in the first place’
- *bilakhir, bilakhirä* ‘finally, in the end’
- *ma* ‘then, when’

Two of these words are listed as also having had non-discourse pragmatic meanings: *bäg”dä* ‘after’ was also used in Tatar as a spatial relator meaning ‘through’ or ‘behind,’ and Arabic *ma*

is widely polysemous in both Arabic and Tatar – it was used to mean not only ‘then’ or ‘when,’ but also as ‘that’ (its root meaning), ‘that which’, ‘which’, and ‘there’.

Most of the interactional performatives borrowed from Arabic are still used in present-day Tatar, but two are now considered archaic:

[4.48]

- *täkbiyä* ‘hi’
- *äblän wä märkbaba* ‘hello’

Russian *privyet* ‘hi’ is the equivalent of *täkbiyä* ‘hi’, and is found code-mixed in Tatar.

Finally, as found with the Russian words code-mixed into Tatar, the greatest number of archaic Arabic discourse-pragmatic loanwords are evaluative in nature. This will be seen to be true for Turkish discourse-pragmatic borrowings into Asia Minor Greek as well. First are general evaluatives, equivalent to Russian *kak* ‘how’:

[4.49]

- *nitän* ‘how? in what way?’
- *nitäkin* ‘how? in what way?’
- *nitäkkäm* ‘how? in what way?’

However, unlike Russian *kak*, which is code-mixed only in its grammaticalized evaluative meaning and never in the more basic sense of ‘in what manner’, these Arabic words were apparently used in both senses, for both evaluation and discussion of manner.

The two positive evaluations are interjections, as are the vast majority of negative evaluations:

[4.50]

- *fäbiha wä nigamä* ‘wonderful, great’
- *mägal’urra wä bilbanzay* ‘With a cry of ura! And with a cry of banzai!’

- *äsäfa* ‘How awful!’
- *dad bidad* ‘Alas! How awful!’
- *häybat* ‘Alas! How awful!’
- *khäef* ‘How awful!’

- *khäifa ki* ‘How awful!’
- *wa wäilä!* ‘How awful!’
- *wa däriga!* ‘Oh, how awful!’
- *däriga* ‘O grief! O pity!’
- *älgayaze billa* ‘God forbid!’
- *näguze billahi* ‘God forbid!’
- *yahu* ‘Oh God!’
- *yaläytä* ‘Oh, if only!’
- *wa!* ‘Akh!’

The Russian interjection of negative evaluation *uzhas!*, loosely glossable as ‘how awful’ is found code-mixed in Tatar and is the approximate equivalent of many of the above interjections. The expression *näguze billahi* ‘God forbid!’ is an Islamic formula (Mäkhmütov et al. 1993: 453), and so was perhaps more religiously marked than other interjections invoking God (for example, the present-day use of the Tatar interjection *Allah böersä* ‘God willing’ (lit. ‘if God commands’) appears to be linked more to the speaker’s level of superstition than level of religious belief).

Two negative evaluations were not interjections, but rather parenthetical adverbs:

[4.51]

- *mä attäässef* ‘unfortunately’
- *täässefän* ‘unfortunately’

Additionally, there are several archaic interjections that are unclear as to whether they would be used for positive or negative evaluation – it is possible to imagine contexts where they would express either stance.

[4.52]

- *älkhökme lilyä* ‘everything is in the hands of God’
- *gajäba* ‘Surprising! How strange!’
- *hya kälyäm* ‘there aren’t words’
- *rabbi* ‘My God!’

While there do not seem to be any archaic Arabic hedges, there were two epistemics:

[4.53]

- *lägallä* ‘maybe, probably’
- *namömkün* ‘impossible, improbable’

The epistemic *lägallä* ‘maybe, probably’ appears to be synonymous with Arabic *bälki* (Mäkhmütov et al. 1993: 241), which is still used in present-day Tatar – therefore it too has a Russian equivalent code-mixed in Tatar, the epistemic adverb *naverno* ‘probably,’ which is used in lieu of *bälki*.

There are seven archaic Arabic borrowings that were used in evaluations of time:

[4.54]

- *däfgatän* ‘immediately’
- *fäürän* ‘immediately’
- *filkehal* ‘immediately’
- *gajälätän* ‘quickly, in a hurry’
- *galälgajlä* ‘quickly, soon’
- *karibän* ‘soon’
- *mötäwazıyan* ‘simultaneously’

The code-mixed Russian *skoro* ‘soon’ is the approximate equivalent of *galälgajlä* and *karibän*, while code-mixed Russian *srazu* ‘immediately’ is the approximate equivalent of *däfgatän*, *fäürän*, and *filkehal*.

Many more of the archaic Arabic discourse-pragmatic borrowings evaluated degree:

[4.55]

- *bäli* ‘yes, true’
- *bilgakes* ‘on the contrary’
- *khasha vä khasha* ‘no-no, never’
- *läysä* negative particle

- *binnisbë* ‘comparatively’
- *kıyasän* ‘comparatively’
- *nisbätän* ‘relatively’
- *käännähü* ‘as if, as though’
- *käma* ‘like, exactly like’

- *äjmägıyn* ‘all’

- *jämig* ‘all’
- *bäg*”*zi* ‘none of...; none’

Only one of these thirteen words, *bäli* ‘yes, true’, has a code-mixed Russian equivalent: *da* ‘yes.’

The final, and largest, category of archaic Arabic evaluations is composed of those words and phrases used to evaluate the relationship between a proposition and reality:

[4.56]

- *bilkbak* ‘in reality’
- *därkbäkijkat* ‘in reality’
- *sadiykana* ‘truly’

- *äbädän* ‘forever’
- *äbädiyän* ‘forever’
- *jawidanä* ‘forever’
- *taäbäd* ‘forever’
- *daimän* ‘always, constantly’
- *häräynnä* ‘always, constantly’

- *äksäri* ‘in the majority of cases’
- *äksäriyä* ‘in the majority of cases’
- *bijömlä* ‘in general’
- *gadätän* ‘usually’
- *gorfän* ‘usually’
- *tgälälgadät* ‘as usual’
- *äke*”*lidän* ‘as usual’
- *jabäja* ‘here, there, and everywhere’
- *mäkhzan* ‘exceptionally, only’

- *bäg*”*zän* ‘sometimes’
- *märrä* ‘one time, once’
- *mirar* ‘repeatedly, more than once’
- *mirarän* ‘several times’

- *galälitlak* ‘absolutely, generally’
- *bishäk* ‘without a doubt’
- *lya mäkhälä* ‘without a doubt’
- *lya shäk* ‘without a doubt’

- *möttäkäyn* ‘without a doubt’
- *bittämam* ‘completely’
- *bitämamiba* ‘completely’
- *bikhakkäy* ‘completely’
- *kyamilän* ‘completely’
- *torra* ‘completely’
- *bänagyah* ‘unexpectedly’
- *galälgaflä* ‘unexpectedly’
- *min khäyse lya yakhtäsib* ‘unexpectedly’

The compound *häraynnä* ‘always, constantly’ is not entirely Arabic, but rather a compound of Arabic *aynnä* and Persian *här*, and is parallel to the mixed compound *härwakit* ‘always’, used adverbially in present-day Tatar and created from Persian *här* and Arabic *wakit* ‘time’ (discussed above in section 4.3.1). Ten of these archaic words have Russian equivalents that are code-mixed today: the code-mixed equivalent of Arabic *biljömlä* ‘in general’ is Russian *v oobshchem*; for *gadätän* ‘usually’ and *gorfän* ‘usually’ there is Russian *obychno*; for *bäg”zän* ‘sometimes’ there is Russian *inogda*; for *galälitlak* in its meaning of ‘absolutely’ there is Russian *absolutno*, while for its meaning of ‘generally’ there is Russian *voobshche*; and for the Arabic borrowings *bittämam* ‘completely’, *bitämamiba* ‘completely’, *bikhakkäy* ‘completely’, *kyamilän* ‘completely’, and *torra* ‘completely’ there is Russian *sousem*.

4.3.4b Archaic Persian borrowings⁷

The subordinating conjunctions borrowed from Persian were:

[4.57]

- *äz baskeh* ‘because’
- *ancheb* ‘that which, that’

⁷ I would like to thank Mino Mo’allah for her help in deciphering the Arabic script in which these words were written in the relevant dictionary.

- *chun, cho* ‘because’
- *ta* ‘so’
- *taakeh* ‘so, in order to’
- *zan, azan* ‘so, from that’

Several of these coordinators have Russian equivalents that are code-mixed in Tatar today: Russian *potomu chto* ‘because’ is the equivalent of *äz baskeh*; and Russian *chtoby* ‘in order to’ is the equivalent of *taakeh*.

There are just two deictic Persian borrowings used for metacommentary that are now archaic in Tatar:

[4.58]

- *in, inha* ‘this, these’
- *bädär* ‘there’

Three positive evaluatives from Persian are no longer used:

[4.59]

- *shabash* ‘bravo!’
- *zebi* ‘bravo!’
- *aya* ‘wonderful! really?’

Two Persian epistemics, both reasonably equivalent to the Russian *mozhet* found in Tatar, were once used:

[4.60]

- *bayäd* ‘maybe, it follows that’
- *shaayäd* ‘maybe, might become’

Finally, the greatest number of now-archaic Persian discourse-pragmatic borrowings evaluated time and degree:

[4.61]

- *ari* ‘yes, like that’
- *bäli* ‘yes, like that’
- *häman* ‘very quickly’
- *baaz* ‘again’

- *hämishä* ‘constantly, continuously’
- *häräy* ‘each time’
- *häräyız* ‘each time’

- *tänha* ‘one, only’
- *chänd* ‘several’
- *asa* ‘like, similar’
- *basaan* ‘like, as if’
- *hänüz* ‘still, also; exactly’
- *bas* ‘sufficiently’
- *bäsa, bäsi* ‘much, enough’
- *bisiyar* ‘much, very’

Two of these words have Russian equivalents code-mixed in Tatar today: while Persian *hämän* ‘very quickly’ is no longer used, Russian *srazu* is, and while Persian *hänüz* is no longer used in the sense of ‘still’, Russian *eshche* ‘still’ can be found.

To conclude, once again we have seen that these now-archaic borrowed discourse-pragmatic words fit into the same three functional categories of discourse structuring words, interactional performatives, and words of evaluation, and many of the Arabic and Persian discourse-pragmatic words that are now archaic have equivalents that are still used to this day. Due to the lack of textual evidence, it is not clear what effect these words had on Tatar morphosyntax, but sheer numbers alone show that Arabic and Persian were quite influential: altogether there have been 224 Arabic and Persian discourse-pragmatic words used in Tatar, whether as nonce borrowings or as standard borrowings, and 54 of these discourse-pragmatic words are still used to this day.

4.3.4c Archaic Arabic and Persian grammatical borrowings

Recall that when the majority of Arabic and Persian words borrowed into Tatar fell into disuse, mostly because of 19th and 20th century lexical reform, all of the non-content borrowings that survived into the present-day had discourse-pragmatic functionality. However, not all of the Arabic and Persian non-content words borrowed into Tatar were discourse-pragmatic words: 23 grammatical words without discourse-pragmatic functionality were also borrowed into Tatar. These borrowings, along with the polysemous use of such words as Arabic *kadär* ‘as much as, up to’ and Persian *häm* ‘and’ in their root meanings or functions (as a spatial relational postposition and grammatical coordinating conjunction, respectively), imply that Tatar was mostly likely at the third stage of lexical borrowing from Arabic and Persian. (Note that the process of a literary language borrowing from a literary superstrate is different from that of a vernacular borrowing from another vernacular – the influence of Arabic/Persian on literary Tatar takes place in a different context from the influence of Slavic on contracting Karaim.) It is interesting to note that these grammatical borrowings disappeared with the waning of Arabic and Persian influence, which suggests that they may not have been as assimilated in the populace as the discourse-pragmatic borrowings, and perhaps were standard borrowings only among highly educated Tatars who were fluent in Arabic and Persian. Their designation as standard borrowings rather than nonce borrowings rests upon the assessment of just a few lexicographers, none of whom give textual evidence – therefore, the placement of the Tatar-Arabic/Persian contact situation within the hierarchical model of borrowing described here awaits further study.

The grammatical non-discourse-pragmatic words borrowed from Arabic and Persian fall into several clear categories. There are spatial relators, pronouns, words that are

combinations of spatial relators and pronouns, and question words. Additionally, many cardinal numbers have dictionary entries, but it is not clear if they were nonce borrowings or consistently used – in any case, they did not supplant any of the native numerals.

Two question words were borrowed into Tatar from Persian:

[4.62]

- *chändan* ‘how many?’
- *cheb* ‘what kind of’

Archaic Arabic and Persian spatial relators in Tatar were:

[4.63]

- *gan* ‘from; to; together; about’ (Arabic)
- *ila* ‘to, up to; for’ (Arabic)
- *fäük* ‘on, higher’ (Arabic)
- *äz* ‘from’ (Persian)

Without textual evidence, it is impossible to know if their use was syntax-altering: for example, whether they were used as postpositions like native Tatar spatial relators, or whether they governed the same cases as their Tatar counterparts.

Four pronouns were borrowed from Arabic and Persian:

[4.64]

- *äna* ‘I’ (Arabic)
- *lya shäi* ‘nothing’ (Arabic)
- *chiz* ‘anything, nothing’ (Persian)
- *käs* ‘anyone, no one’ (Persian)

It seems highly unlikely that this list is complete – for example, what would motivate the borrowing of only one personal pronoun? Even so, these are the only pronouns that had entries in all the relevant dictionaries.

Also borrowed from Arabic were words that compound spatial relators and pronouns, mostly using the spatial relator *gan* in its meaning of ‘from’:

[4.65]

- *gankem* ‘from you; with you; about you’
- *ganna* ‘from us’
- *ganba* ‘from her; about her’
- *ganbe* ‘from him; about him’
- *ganbem* ‘from them; about them’
- *ganbema* ‘from them both; about them both’
- *läb* ‘to him, for him’

From Arabic and Persian we find two words meaning ‘for, for the sake of’:

[4.66]

- *läjäl* ‘for, for the sake of’ (Arabic)
- *baraye* ‘for the sake of’ (Persian)

Finally, we find several adverbs of manner:

[4.67]

- *mäsürän* ‘happily’
- *mötäwazigan* ‘humbly, quietly’
- *mötägammidän* ‘seriously, assiduously’
- *möshfikanä* ‘finely, gently’

All other adverbs of manner borrowed from Arabic into Tatar, for example *makhsus* ‘especially’ or *gomumän* ‘generally’, have clear discourse-pragmatic functionality. However, it is difficult to imagine a context where the four adverbs listed above could be used as either metacommentary or as some sort of evaluation – they do not appear to express stance, only manner. This expansion in borrowing from only adverbs of stance to include adverbs of manner will be addressed further in the final section of this chapter.

In conclusion, we have seen that the present-day phenomenon of code-mixed Russian words in otherwise Tatar speech is not the first time that discourse-pragmatic words have been borrowed into Tatar. In fact, while Tatar appears to be only just entering the second stage of lexical borrowing in its contact with Russian, the evidence presented here suggests that literary Tatar was, at one point, in the third stage of lexical borrowing from

Arabic and Persian, borrowing not only content words and discourse-pragmatic words, but also grammatical and function words. However, when the wave of Arabic and Persian influence receded, only the content-word and discourse-pragmatic borrowings remained behind. The influence of a literary superstrate on a literary language is clearly different than the influence of a dominant vernacular on a contracting minority language; the ways in which this is so invite further study.

4.4 Case Study Three: Turkish influence on Asia Minor Greek

Thomason and Kaufman (1988) used the case of Asia Minor Greek under the influence of Turkish as the exemplification of a language with the very heaviest level of lexical and structural borrowing, located at the extreme end of the borrowing scale (level 5). In this section, I will use the same source of data as Thomason and Kaufman, Dawkins' (1916) grammatical analysis of the Greek dialects of Asia Minor. Using data taken from the hundreds of pages of folktales collected by Dawkins, I will show that a significant amount of the lexical and structural borrowing of Asia Minor Greek is best understood by using a model where discourse-pragmatic functionality is an organizing principle.

Dawkins collected his dialectal data and texts just a few years before the expulsion of Asia Minor Greeks from Turkey in the 1910s, an expulsion that significantly reduced the number of speakers of these dialects. Even so, the Greek dialects of Asia Minor had been contracting for hundreds of years and continued contracting after the population exchange: the earliest evidence of language shift dates from 1437 (Dawkins 1916: 1), and between 1927 and 1955 the number of Greek speakers recorded in the Turkish census dropped by approximately 40,000 (Thomason and Kaufman 1988: 215). Turkish speakers do not appear

to have shifted to Greek, thus discounting substrate influence (ibid. 216). Dawkins characterized the Greek of Cappadocia (the most heavily influenced dialect region) as a language where “the body has remained Greek, but the soul has become Turkish” (1916: 198). In this section, I will show that the manner in which the soul of Asia Minor Greek became Turkish is clearly related to the manner in which Tatar has been and is being influenced by Russian, Arabic and Persian, and in which Karaim has been influenced by Russian and Polish: discourse-pragmatics is key, and the case of Asia Minor Greek will provide further evidence that it is the same discourse-pragmatic words (or words with similar meanings and functions) that are borrowed again and again in similar language contact situations. It was shown in the previous section that many of the “native” Tatar words in the process of being replaced by code-mixed Russian discourse-pragmatic words are themselves borrowings from Arabic or Persian – one wave of discourse-pragmatic borrowings in Tatar is replacing words from an earlier wave of discourse-pragmatic borrowings. An examination of the 47 Turkish discourse-pragmatic words borrowed into Asia Minor Greek shows a similar phenomenon, but here instead of discourse-pragmatic loanwords from one language replacing synonymous discourse-pragmatic loanwords from another language, we find the same discourse-pragmatic words borrowed twice – first from Arabic or Persian into Turkish, and then later from Turkish into Greek. Eight of the Turkish discourse-pragmatic loanwords or phrases found in Asia Minor Greek were originally borrowed into Turkish from Persian, and fifteen of the words and phrases were partially or entirely borrowed from Arabic – in other words, 49% of the Turkish discourse-pragmatic words borrowed into Asia Minor Greek are twice-borrowed words, borrowed once from a highly influential literary superstrate into Turkish, and then borrowed again from a dominant local vernacular into contracting Asia Minor Greek. In addition, all but one – or 96% – of these twice-borrowed

Arabic and Persian discourse-pragmatic words are also found as loanwords in Tatar.⁸ Appendix E of this dissertation summarizes the Turkish discourse-pragmatic words, their etymologies, and their Tatar counterparts so that this information can be seen at a glance.

4.4.1 Turkish discourse-pragmatic borrowings in Asia Minor Greek

The forty-seven discourse-pragmatic words borrowed from Turkish into Asia Minor Greek presented here come from Dawkins (1916), and were culled both from the glossary and the collected folktales. Excerpts from the folktales used as illustrations of the discourse-pragmatic words in context all come from Dawkins (1916), and are cited with both the page number and dialect (as Asia Minor Greek was not monolithic, and there were varying degrees of lexical and structural borrowing in the various dialects). When discussed in the text, Turkish words are presented in the pan-Turkic orthography used for both Tatar and Karaim elsewhere in this dissertation (see Appendix A for notes on transcription); folktale excerpts are written in IPA and were converted from the modified Greek alphabet used by Dawkins for broad phonetic transcription. Therefore, the Turkish borrowings found in the Greek excerpts in the examples will usually be spelled differently than in the surrounding text – however, their phonetic realizations in Greek, if different, are found in the bulleted lists that begin each subsection. All Turkish borrowings, whether discourse-pragmatic terms or content-word borrowings, are shown in boldface so that both the frequency and the placement of all borrowings can be assessed with ease.

⁸ There is no way to conclusively state that the Arabic-Persian contact situations for Turkish and Tatar were entirely separate and that Tatar and Turkish independently borrowed the Arabic and Persian discourse-pragmatic words that they have in common. However, given historical circumstances, independent borrowing seems the most likely scenario.

Turkish discourse-pragmatic borrowings in Greek, like all the discourse-pragmatic borrowings discussed to this point, fall into three now-familiar categories: (1) Markers of discourse structure and force, including subordinative, coordinative and contrastive discourse structuring words, as well as words used for the purposes of metacommentary and deixis; (2) Interactional performatives, and (3) Evaluatives, including words used as positive evaluation, as epistemics and deontics, and to evaluative relative time and degree.

Discourse-structuring subordinate conjunctions

There are five Turkish subordinating conjunctions found in the dialects of Asia Minor Greek:

[4.68]

- *ki* ‘that’
- *chünki* ‘because’, in Greek as [tʃynki], [tʃunki]
- *ichin* ‘for, on account of, because’, in Greek as [itʃin], [itʃyn].
- *mademki* ‘because, as long as’
- *eyer* ‘if’, in Greek as [éjer], [er].

Three of these conjunctions – *ki*, *chünki*, and *eyer* – are themselves borrowings from Persian, and have equivalents – *ki*, *chönki*, and *ägär* – that are used in Tatar to this day (see section 4.3.2 above). The first of these, *ki* ‘that’, is used in Asia Minor Greek not only to introduce subordinate clauses, but also to introduce both indirectly and directly reported speech, as seen in example [4.69]:

[4.69] Turkish *ki* used to introduce direct speech

sónadan rótse ki, “pú pénis.”
afterwards ask.AOR-3PS that where go.PRES-2PS⁹
‘Afterwards he asked him, “Where are you going?”’ (Ulaghátsh dialect: 366)

While in Turkish *ki* can introduce both subordinate clauses and indirect speech, it is not used to introduce direct speech¹⁰ – this function appears to be an innovation in the borrowing Greek dialects. *Ki* was also combined with Greek *ton* ‘when’ to form another subordinating conjunctive, *tonki* ‘when’, seen in example [4.70]:

[4.70] Innovated Turkish-Greek *tonki*

ke tóngi pan enjá mínes
and when go-SIMPPST.3PS nine months
‘...and when nine months pass...’ (Delmesó dialect: 316)

Example [4.71] has both *ki*, used once again to introduce direct speech, and the Turkish subordinating conjunction *chönki* ‘because’:

[4.71] Turkish *ki* and *chönki*

ekinó épe ki, “xabár dén éjijga,
that one say-SIMPPST.3PS that news NEG touch.SIMPPST-1PS
tşyñki jípnoşa.”
because sleep.AOR-1PS

He said, “I have no **knowledge**, **because** I went to sleep.” (Ulaghátsh dialect: 348)

⁹ Dawkins gives only free translations of the texts, and thus while the free translations in this section come directly in this section come directly from Dawkins (1916), the interlinear glosses are my own. A key to the grammatical abbreviations in the glosses can be found in Appendix A. I would like to thank Maria Kotzamanidou and Andrew Garrett for their help with some of these glosses, particularly with regards to phonetically reduced and altered dialectal variants of grammatical words. Any and all mistakes are the sole responsibility of the author.

¹⁰ Many thanks to both Ayla Algar and Karl Zimmer for their consultation on Turkish grammar. Once again, any mistakes in interpretation are the sole responsibility of the author.

The subordinating conjunction *chünki* ‘because’ was also borrowed from Persian into Tatar as *chönki*, where it is used to introduce subordinate clauses while expressing causality, just as it does in Turkish and as shown here in Asia Minor Greek. Adding one more level to this web of parallel and repeated borrowing of discourse-pragmatic words is the fact that Russian *potomu chto* ‘because’ is now sometimes code-mixed in Tatar in lieu of *chönki*, and may be en route to replacing it as the subordinating conjunctive of choice.

While subordinating conjunctions borrowed from Persian into Turkish into Greek introduce subordinate clauses that follow main clauses, as one might expect,¹¹ the Turkish subordinating conjunction *ichin* ‘on account of, because’, when borrowed into Greek, is used in Turkic-style pre-head embedded subordinate clauses, as shown in example [4.72]:

[4.72] Turkish *ichin* ‘on account of, because’ in a pre-head clause

...ge épe **ki** áp to **sevdú-si** **itşin**, “Deré vavá m
and say **that** from the **love-POSS** **on account of** now father my

na ért, je ná se rotíş, ‘**baqalim** emaéz **mi,**’ ” dí.
FUT come and FUT you ask **let’s see** learn.PRES-2PS ?PT say.PRES-3PS

‘...and **because** of her **love** she said, “now my father will come and ask you, **let us see**, have you learned anything?’ ” ’ (Ulaghátsh: 366)

There are many interesting instances of lexical and morphosyntactic borrowing in this one example. Here we see *ichin* in its Turkic phrase-final position in a subordinate clause that precedes the main clause – compare this to the post-head clause introduced by (originally Indo-European) *chönki* seen in example [4.71] above. Turkish *ki*, used once again to quote direct speech, is no longer adjacent to the direct quote, as it is in examples [4.69] and [4.71] –

¹¹ The borrowing of Persian *chünki* appears to have been syntax-altering in Turkish, just as it was in Tatar.

this is because the phrase-final slot is taken by *ichin*. Two additional discourse-pragmatic borrowings are also found in the example: the interrogative particle *mi*, found utterance-finally (in the quoted speech), just as it would be in Turkish, and the optative form of the Turkish verb *bakmak* ‘see’, used here as a discourse directive bringing attention to the information that is about to be presented. In addition to Turkish discourse-pragmatic borrowings and the more Turkic syntax associated with their use, there are several other instances of Turkish influence in this example. The first is the use of a Turkish possessive suffix on *sevdu* ‘love’ to mean ‘her love’ – this use of Turkish morphology to express possession appears to have supplemented rather than entirely replaced Greek morphology, which will be seen later in this section. Note that the introduction of the Turkish possessive suffix $-(s)I$ into an Indo-European language is the reverse of the lessened use of this suffix in the *izafet* constructions found in Turkic Karaim under the influence of Slavic, discussed in section 4.2 above. Additionally, the word *vavá* ‘father’ is also possessed Turkish-style: the Greek first-person possessive pronoun is phonetically reduced to just *m* and placed after the modified noun such that it is indistinguishable from the Turkish first-person singular possessive suffix $-(I)m$ – compare *vavá m* ‘my father’ as found here to Turkish *baba-m* ‘my father’. Also nearly homonymous with its Turkish equivalent and used identically is *di*, the third-person singular present tense of the Greek verb ‘to say’ – placed here right after the directly quoted speech, just as Turkish gerundial *diye* ‘saying’ would be. In fact, echoing Turkish word order, most of the verbs in this example, as in many other examples found in this section, are phrase- and clause-final.

Turkish *mademki* ‘since, as’, borrowed into Silli Greek with a meaning more along the lines of ‘because’, appears to be a compound formed by Turkish *madem* ‘since, as’ and Persian *ki*, much like the compound Greek-Turkish(/Persian) *tonki* ‘when’ seen above in

example [4.70]. Perhaps it is due to this Persian element that *mademki* is used clause-initially to introduce a subordinate clause, as seen in example [4.73]:

[4.73] Turkish *mademki* ‘since, while’

madémki	ʃí	oz	zariá	xitʃ	pará	rén	iripsis...
because	you.sg	until	now	nothing	para	not	seek.contpast-2PS

‘Because until now you have never asked for **any** money’ (Silli dialect: 292)

Turkish *hiç* ‘nothing, none’, realized here as [xitʃ], requires a verb in the negative – the Greek verb in the example here is in the negative, and thus congruent with the Turkic syntactic requirement, but it should be noted that the Greek equivalent also requires a verb in the negative, so this can not be seen as an instance of structural influence. Turkish *hiç*, like Tatar *hiç*, is itself a borrowing from Persian.

Also originally a borrowing from Persian is the final Turkish subordinating conjunction found in Asia Minor Greek, *eyer* ‘if’ (borrowed into Tatar as *ägär*), found in example [4.74]:

[4.74] Turkish *eyer* ‘if’

Ejer	na	páres	tutunu	tʃiŋ	góri	oz	enéka,
if	FUT	take-2PS	this one.GEN	the	girl	as	wife

tóti	ʃí	se	na	nis	polí	zengínis.
then	you	FUT	SUBJ	become.2PS	very	rich

“If you take this man’s daughter as wife, then you shall become very **rich**.” (Silli dialect: 288)

Just as in Turkish (and in Tatar, for that matter), *eyer* is sentence-initial when used in Silli Greek, but here it is used as the sole indicator of conditionality, whereas in Turkish (and in Tatar), it is optional and used in addition to a verb with the conditional affix -ɛA-.

Coordinating discourse markers

Asia Minor Greek borrowed three coordinating discourse markers from Turkish:

[4.75]

- *hem* ‘and, also’, in Greek as [xem]
- *da* ‘and’ (enclitic), in Greek as [da] or [de]
- *demek* ‘so, thus’

The first of these coordinators, *hem* ‘and’ is a Persian loanword in Turkish – the Tatar equivalent *häm* ‘and’ was discussed above in section 4.3.3. An example of Turkish *hem* in Asia Minor Greek is shown in [4.76]:

[4.76] Turkish *hem* ‘and’

Ke	óson k	íðen	to	peđí,	tó ná to	ftʃif	son dópo,
and	as soon as	saw	her	boy	the act of	spit	instead of
xém	ðéken	do k	éfayen,	xém	filsen		džin.
and	give-SIMPPST-3PS	it	and eat-SIMPPST-3PS	and	kiss.SIMPPST-3PS		her

‘And as soon as the boy saw her, instead of spitting at her, he **both** gave her food and she ate, **and** he kissed her.’ (Delmesó dialect: 324)

Note that Greek *kai* ‘and’ (and its reduced variant *k*) can be used both to link objects and to create narrative structure. However, Turkish *hem*, much like Russian *i* ‘and’ when code-mixed in Tatar, is used only for narrative structure. Persian *häm* is used in present-day Tatar in both functions, as a general coordinating conjunction and to structure discourse, and this use of *häm* in Greek only for narrative structure gives further evidence that polysemous (or multi-functional) words are borrowed first in their discourse-pragmatic function and only later used as grammatical or function words without discourse-structuring or pragmatic force. This also suggests that there was a time when Persian *häm* in Tatar was used only for

discourse structuring; perhaps with further research, texts can be located that show that this was, indeed, the case.

Like *häm*, the Turkish enclitic *da* ‘and’ is also used only to structure narrative when used in Greek, as shown in example [4.77]:

[4.77] Turkish *da* ‘and’

ape	dío	tréyne,	pén	de	‘s	to	líka	embró,
DIST	two	run.PRES-3PP	go.PRES-3PP	and	to	the	wolf	in front
ke	líko	trói	ta					
and	wolf	eat.PRES-3PS	them					

‘Two by two they run, **and** go up to the wolf, and the wolf eats them.’ (Axó dialect: 398)

As an enclitic, Turkish *da* assimilates in vowel quality to the preceding word and is realized as *de* after words with front vowels, and it seems possible that it retained its sensitivity to the rules of Turkish vowel harmony even when borrowed into Greek – here it is realized as *de* after the verb *pén* ‘he goes’, which has a front vowel. Note that it has also retained its Turkic post-head word order, and does not introduce the clause but rather follows the verb that it is coordinating.

Contrastive discourse markers

With the exception of *yoksa* ‘or’, the Turkish contrastive discourse markers borrowed into Asia Minor Greek are loanwords that were originally borrowed from Persian and Arabic:

[4.78]

- *meyer* ‘but’
- *amma* ‘but’, in Greek as [ammá], [amá], [má]
- *lakín* ‘but’
- *yabut* ‘or’, in Greek as [yaxút]
- *yoksa* ‘or’, in Greek as [yúsa]

Meyer ‘but’ and *yabut* ‘or’ were borrowed from Persian into both Turkish and Tatar, and *amma* ‘but’ and *lakin* ‘but’ were borrowed from Arabic into both Turkish and Tatar. Additionally, these four words have Russian equivalents that are code-mixed in present-day Tatar: Russian *no* ‘but, however’ is used in lieu of *meyer*, *amma*, and *lakin*, while Russian *ili* ‘or’ is used in lieu of *yabut*. An example of *meyer* in context can be seen in [4.79]:

[4.79] Turkish *meyer* ‘but’

Mejér	to	şafki	tu	Duniá	Gyzel-i-đju	tun,
but	the	brightness	the.GEN	world	beauty-POSS-GEN	the.GEN
tú		kaθótun	so	miván	apáno.	
be.SIMPPST-3PS		sit-CONTPST-3PS	in	fruit tree	on	

“**But** the brightness was that of the **Fair One of the World**, who was sitting up in the **fruit tree**.” (Delmesó dialect: 310)

As in Tatar and Turkish, *meyer* is found sentence-initially and clause-initially when borrowed into Greek. Also of interest in this example is the use of Greek inflectional morphology on a borrowed Turkish phrase – the genitive *-đju* found on *Duniá Gyzeli* ‘Beauty of the World’ (translated by Dawkins as ‘Fair One of the World’), itself a compound noun itself created by means of Turkish *izafet* (the *-i* suffixed on *Gyzel* ‘beauty’ is what compounds it with *Duniá* ‘world’) and apparently borrowed as a unit. This is in contrast to *sevdu-si* ‘her love’, as seen in example [4.72] above, which seems to show sensitivity to the rules of Turkish inflectional morphology.

A final example of contrastive markers shows Turkish *amma* in its Greek context:

[4.80] Turkish *amma* ‘but’

Ammá os pú na nárti, kóre xalánni tu **odzáqi.**
but as where FUT come girl destroy the **hearth**
“**But** before he comes, the girl breaks down the **pyre.**” (Silli dialect: 284)

Like *meyer* ‘but’ and *lakin* ‘but’, *amma* ‘but’ too is always clause- or sentence-initial in Asia Minor Greek, just as it is in Turkish and Tatar.

Metacommentary and deixis

As seen with the discourse-pragmatic borrowings found in the other language contact situations presented in this chapter, the most extensive set of borrowings used to structure or mark discourse are words used for metacommentary and deixis. Thirteen Turkish discourse-pragmatic words and phrases used for metacommentary were borrowed into Greek:

[4.81]

- *mi* interrogative particle, in Greek as [mi] and [mu]
- *o* ‘that’, in Greek as [ó] and [ón]
- *bu* ‘this’
- *sonra* ‘afterwards’, in Greek as [songrá]
- *sonradan* ‘afterwards’ (dialectal), in Greek as [songradán] and [songadán]
- *en bash* ‘first, in the beginning’, in Greek as [em baʃ]
- *bir vakit(ta)* ‘once (upon a time)’
- *yeniden* ‘anew, afresh’
- *mahsus* ‘expressly, in particular’, in Greek as [másustan]
- *filan* ‘and so on, such and such’, in Greek as [filáni] and [filán]
- *kesbke* ‘if only’, used to introduce wishes
- *hayde* ‘come, c’mon’, in Greek as [xájde] and [xáde]
- *bakalım* ‘let’s see’

The interrogative particle *mi*, like the enclitic *da* discussed above, remains subject to the rules of Turkish vowel harmony and is found as *mu* when preceded by words with back vowels,

even if those words are Greek. Let's revisit example [4.72], excerpted here as [4.82] to focus on the metacommentary words:

[4.82] Turkish *mi* (question particle) and *bakalim* 'let's see'

...ge ná se rotíʃ, 'baqalim emaez mi,'
...and FUT you ask.2PS **let's see** learn.PRES-2PS ?PT
'...and ask you, "**let us see**, have you learned anything?"' (Ulaghátsh dialect: 366)

Here the quoted speech (which is not actually a quote, but rather what the magician's daughter predicts her father will say to her young lover) consists of three words, two of which are Turkish metacommentary words. The optative *bakalim* 'let's see' is used as a directive, just as in Turkish, and so was perhaps borrowed holistically as a unit meaning 'let's see' rather than with sensitivity to Turkish morphology. The interrogative particle *mi*, however, appears to remain consistent with Turkish vowel harmony and syntactic requirements: it is here in its unrounded form, as it is preceded by a (Greek) word with an unrounded vowel, and it is placed both utterance-finally and directly after the verb that is in question, just as it would be in Turkish.

Two deictic demonstrative pronouns were borrowed from Turkish into Asia Minor Greek: *o* 'that' and *bu* 'this'. *O* 'that' appears to be the less-widely used of the two, although it does combine with another borrowed discourse-pragmatic word to form a new discourse-pragmatic word, as will be seen later in this section. The demonstrative pronoun *bu* 'this' is used with Turkish words, often words of time, thus creating entirely Turkish phrases used to structure the narrative, as seen in example [4.83]:

[4.83] Turkish *bu* ‘this’

bu	sefér	jémosen	to	fési	t	neró,	k	eθekén
this	time	fill.SIMPPST-3PS	the	fez	his	water	and	put-SIMPPST-3PS

do	kondá	t.
it	near	her

‘**This time** he filled his fez with water, and put it down near.’ (Delmesó dialect: 308)

Here, *bu* is being used to emphasize the contrast between this time and the last time – we are at a point in the folktale where the young hero meets his second Beauty of the World, who has come out of a fruit tree literally dying of thirst. The first time this happened, he did not succeed in bringing a fezful of water sufficiently quickly, and his prospective bride died. Here the *bu* emphasizes that *this* time, although the circumstances are the same, the outcome is different. Turkish *sefer* ‘time’ is used in another narrative structuring expression, where it combines with Greek *ena* ‘one’ to form the phrase [ena sevér] ‘once’.

Many temporal adverbs were borrowed from Turkish into Greek, and the borrowed discourse-pragmatic word most frequently found in the Greek folktales cited here is *sonra* ‘after’ (and its dialectal variant *sonradan*), used for narrative structure as a sort of general continuation, much like English *then* and *and then*. Often *sonra* will be the only borrowed discourse-pragmatic word used in an entire folktale, and it is the only borrowed discourse-pragmatic word found in example [4.84]:

[4.84] Turkish *sonra* ‘after, later’

Songrá	tútus	bénni	kí	t	xurjó,	kját
afterwards	this	enter.PRES-3PS	there	the	village	REL.PRO
tu	ípi		xizyris.			
that.GEN	say.SIMPPST.3PS		holy man			

“**Afterwards** the youth enters that village, of which the **holy man** had told him. (Sílli dialect: 288)

As with so many other discourse-pragmatic borrowings, *sonra* is always phrase- or sentence-initial. Other temporal borrowings used for narrative structure are also found only in initial positions, as one would expect from their discourse-structuring function. In example [4.85] we see another temporal Turkish phrase:

[4.85] Turkish *en bash* ‘in the beginning’

em	báf	to	patifáx	épe	ki,
SUPER	beginning	the	king	say.SIMPPST.3PS	that

‘**In the beginning** the king said...’ (Ulagátsh dialect: 358)

The Turkish superlative particle *en* is only found in this construction, and this, combined with its assimilation in place to the following bilabial consonant, suggests that *en bash*, a set phrase in Turkish, was borrowed into Greek as a single unit. Another temporal phrase that seems to have been borrowed as a single discourse-pragmatic unit is *bir vakit* ‘once’, literally ‘one time’ (note that *vakit* is itself a borrowing into Turkish from Arabic):

[4.86] Turkish *bir vakit* ‘once’

bír	vaqít	kejótan	éna	néka.
one	time	exist.CONTPST.3PS	a	woman

‘**Once** there was a woman.’ (Axó dialect: 390)

This phrase is perhaps better translated as ‘once upon a time’, as it begins a folktale. Adding to the evidence that *bir vakit* was borrowed as a unit is the fact that it is used in the locative case to begin a different folktale in the same dialect:

[4.87] Turkish *bir vakitta* ‘once’

bir	vaqıt-da	kejótan	éna	néka.
one	time-LOC	exist.CONTPST.3PS	a	woman

‘Once upon a time there was a woman.’ (Axó dialect: 398)

It is possible that folktales in the Turkish dialects surrounding these pockets of Asia Minor Greek began with the set phrase *bir vakit(ta)* ‘once upon a time,’ but standard Turkish folktales begin with other temporal contextualizers, such as *bir zamanlar* ‘once upon a time,’ lit. ‘one time-PL’.

Another temporal contextualizer borrowed from Turkish is *yeniden* ‘anew, afresh’, seen in example [4.88]:

[4.88] Turkish *yeniden* ‘anew’

adží	o	vasilós	yéniden	pedžen	da	ýámos
there	the	king	anew	make.SIMPPST-3PS	the	marriage

oftá iméres, oftá niéxtes.
seven days seven nights

‘The king made them a wedding **afresh** for seven days, seven nights.’ (Phárasa dialect: 500)

The other Turkish discourse-pragmatic words borrowed for use as metacommentary are not temporal in nature. Turkish *filan* ‘such and such’, originally from Arabic and also found in literary Tatar (where it is nearly archaic), is used to refer to an unspecified noun, as can be seen in example [4.89]:

[4.89] Turkish *filan* ‘such and such’

Kinde so **filán** son dópo
be.PRES-3PP in the **such and such** in the place
“They are in **such and such** a place.” (Delmesó dialect: 304)

The particle *kesbke*, also from Arabic and also found in archaic Tatar, is found in Asia Minor Greek in clause-initial position, introducing a wish, just as it does in Turkish:

[4.90] Turkish *kesbke* ‘if only’

kéŷge tŷiŷ góri mu mí tŷi skótisa...
if only the girl my not her kill.AOR-3PS
‘**Would that** I had not killed my daughter...’ (Silli dialect: 286)

The final Turkish word used to structure discourse is the exhortative *hayde* ‘come, c’mon’, a verb in the imperative.

[4.91] Turkish *hayde* ‘come’

Xáde, fér do mikró s to koríŷ.
come bring the young your the daughter
‘**Come**, bring your youngest daughter!’ (Ulaghátsh dialect: 374)

As might be expected, *hayde* is only found in the folktales in directly quoted speech, as it is an imperative. This sort of exhortative imperative lends itself to borrowing: Russian *davai(te)* ‘come, c’mon’ is regularly code-mixed in Tatar-preferred style. Additionally, Tatar *äydä* ‘come, c’mon’ has been borrowed into Russian as the exhortative *ayda* ‘come, let’s go’. In Tatar *äydä* is almost always used as a verbal inducement of physical motion, while its borrowed version in Russian is used as a more general exhortative, just as borrowed Russian *davai(te)* is also used as a general exhortative in Tatar.

Interactional performatives

While there are four Russian interactional performatives code-mixed in Tatar (for ‘hi’, ‘bye’, ‘please’, and ‘excuse me’), only one Turkish interactional performative is found in the Greek of Asia Minor: *selam* ‘hello’, itself a borrowing from Arabic (and found in Tatar). This is most likely due to the nature of the discourse used in this analysis – it is all narrative, and only in folk-tale register. If Dawkins had also recorded conversational speech, which is where these interactional performatives are used, presumably we would find additional Turkish borrowings used in this function. *Selam* can be seen in example [4.92]:

[4.92] Turkish *selam* ‘hello’

Se tu róis **selámi.**
You him say.PRES-2PS **hello.**
“You will **greet** him.” (Silli dialect: 288)
More literally, ‘You will say “**hello**”.’

Note that this is not actually an instance of *selam* being used as an interactional performative as might be found in quoted dialogue, where one person would greet another by saying “*Selam*” ‘Hello’, but rather an allusion to the fact that *selam* is used as a greeting.

Evaluatives

As with interactional performatives, there are fewer Turkish evaluatives found as loanwords in the Greek folktales cited here than one might expect. Once again, this is most likely due to the nature of the corpus, composed entirely of narratives; in a corpus that also had conversational data, one might expect to find motivation for the use of additional borrowed evaluatives. There are no Turkish words used as negative evaluations or as hedges – for

comparative purposes, recall that Russian negative evaluators meaning ‘How awful’ and ‘It’s a shame’ are found code-mixed in Tatar, and that Russian hedges meaning ‘only,’ ‘just’, and ‘about, around’ are also found code-mixed in Tatar. What we do find are one word of positive evaluation, epistemics, and evaluators of time and degree.

The one Turkish word of positive evaluation found in Asia Minor Greek, *aferin* ‘bravo, well done’ (in Greek as [aferim]), is a borrowing from Arabic that is also found in Tatar. Additionally, the Russian equivalent, *molodyets* ‘bravo, well done’ is frequently code-mixed in Tatar to express approval.

There are two Turkish epistemics that were borrowed into Asia Minor Greek:

[4.93]

- *belki* ‘perhaps, probably’, in Greek as [bélki] and [bérki]
- *mutlaka* ‘without a doubt, absolutely’, in Greek as [nútlaka]

By now it should come as no surprise that both *belki* and *mutlaka* are twice-borrowed words, first coming into Turkish from Arabic, and then into Greek from Turkish. Additionally, they are both found in present-day Tatar, although Tatar *mutlakan* is nearly archaic. A typical use of *belki* is found in example [4.94]:

[4.94] Turkish *belki* ‘perhaps, probably’

bélki	‘a	návri	ti	qoláin du
perhaps	SUBJ	find-3PS	the	solution

“Perhaps he will find what is to be done.” (Pharása dialect: 538)

Turkish *mutlaka*, while used in Turkish both as an epistemic and as an evaluator of a proposition’s relationship to reality (meaning ‘absolutely’), seems to have been used in Asia Minor Greek only as an epistemic, a usage that can be seen in example [4.95]:

[4.95] Turkish *mutlaka* ‘without a doubt’

efis fólki klóθete, to pú índe **nútlaka**
you.PL while walk about.PRES-2PP the where be.PRES-3PP **without doubt**

ksévreté to.
know.PRES-2PP it

‘Since you move about, **without doubt** you know where they are.’ (Delmesó dialect: 306)

In Tatar the word *motlakán*, when still used, is not found in this meaning, but only as an evaluative adverb meaning ‘absolutely’.

Evaluation of time, degree, and relation to reality

Asia Minor Greek borrowed three Turkish evaluators of time:

[4.96]

- *birden bire* ‘at once, immediately’
- *daba* ‘still; already’, in Greek as [daxá]
- *bele* ‘even, already’

All three of these words have Russian equivalents that are code-mixed in Tatar: *srazu* ‘immediately’ and *uzbe* ‘already’. A typical example of *birden bire* ‘at once’ is seen in example

[4.97]:

[4.97] Turkish *birden bire* ‘at once’

“**birdén biré** na fígo” dí.
at once FUT flee-1PS say.PRES-3PS
‘“**At once** I will flee away” he said.’ (Axó dialect: 394)

More frequently borrowed from Turkish were evaluators of degree:

[4.98]

- *pek* ‘very’
- *tamamen* ‘completely’, in Greek as [tamán]
- *daba* ‘still’, in Greek as [daxá]
- *beraber* ‘equal, uniform’, in Greek as [barabári]
- *dek* ‘as far as, until’, ‘like, similar to’
- *kadar* ‘as much as, as many as’ In Greek as [qadar]
- *o kadar* ‘as much as that, so much’, in Greek as [óadar]

Two of these words, *tamamen* ‘completely’ and *kadar* ‘as much as’ are twice-borrowed Arabic words that are also found in Tatar, and Russian *sousem* ‘completely’, the equivalent of *tamamen*, is found code-mixed in present-day Tatar. An example of the standard Turkish intensifier *pek* can be seen in [4.99]:

[4.99] Turkish *pek* ‘very’

raxátjín du **pék** polí kaló iton
pleasure his **very** very good be-PST
‘His **pleasure** was **very** great.’ (Silli dialect: 290)

Note that here *pek* is being used for double-intensification: Greek *kaló* ‘good’ is already modified by the Greek intensifier *polí* ‘very’.

The Turkish discourse-pragmatic word *daba* ‘still; already,’ is used both as a temporal evaluator and as an intensifier that forms a comparative – its comparative use is shown in example [4.100]:

[4.100] Turkish *daba* ‘still; already’

ke ás ála ‘sa dió **daxá** **gyzél** ton.
and from other.PL the two **still** **pretty** be.SIMPPST-3PS
‘...and she was **still fairer** than the other two.’ (Delmesó dialect: 311)

Other Turkish evaluators such as *dek*, *beraber*, and *kadar* are used to show an equality of degree: *beraber* ‘equal, uniform’ and *kadar* ‘as much as’ are shown in examples [4.101] and [4.102] respectively:

[4.101] Turkish *beraber* ‘equal, uniform’

ke eǵenan m  to b y m barab ri
 and become with the height my equal
 ‘...and become **as tall as** I am’ (Delmes  dialect: 318)

[4.102] Turkish *kadar* ‘as much as’

Otji qadar l res ke  n  elis,   se d skum.
 they **as many as** lira and if wish.PRES-2PS, there you give.PRES-1PP
 ‘**As many** pounds **as** you want, we will give you.’ (Delmes  dialect: 316)

In Turkish, the demonstrative pronoun *o* ‘that’ combines with *kadar* to form the phrase *o kadar* ‘as much as that, so much’ (Turkish also has the phrase *bu kadar* ‘as much as this, so much’, but this phrase is not found in the available texts). The phrase *o kadar* ‘so much’ is found in Greek as the phonetically reduced [ adar], as seen in example [4.103] below – it is not clear if this phonetic reduction took place in the local Turkish dialects or in Greek after the phrase was borrowed.

[4.103] Turkish *o kadar* ‘so much, as much as that’

edeka  adar par ja
 give.SIMPPST-1PS **so much** para.PL
 ‘I gave **so much** money...’ (Ulagh tsh dialect: 368)

In addition to evaluators of time and degree, speakers of Asia Minor Greek borrowed Turkish words that evaluate the relationship of a proposition to reality:

[4.104]

- *eyi* ‘yes, very well’, in Greek as [ilíki], [elík], and [ex]
- *hayir* ‘no’, in Greek as [khaír] and [khaíri]
- *hich* ‘nothing’, in Greek as [xetʃ]
- *her* ‘every’, in Greek as [xér]
- *tı̇pkı̇* ‘quite like, similar to’

Many of these words are twice-borrowed discourse-pragmatic words of Arabic or Persian origin, have a Russian equivalent that is code-mixed in Tatar, or both. Russian *da* ‘yes’ is found code-mixed in Tatar, and is the equivalent of Turkish *eyi*, while Russian *nyet* ‘no’, also found code-mixed in Tatar, is the equivalent of Turkish *hayir* ‘no’, which is itself originally from Arabic. Turkish *tı̇pkı̇* ‘quite like, similar to’ is also of Arabic origin – it is the only Arabic or Persian discourse-pragmatic term borrowed into Turkish and then borrowed again into Asia Minor Greek that is not also found in Tatar, although the nominal form of its trilateral root was borrowed into Tatar as *tı̇ybk* ‘similarity, conformity’, a word that is now archaic. Finally, Turkish *hich* ‘no, none’ and *her* ‘each’ are borrowings from Persian, also found in present-day Tatar.

A typical example of borrowed Turkish *hayir* ‘no’, used to negate a proposition, is seen in [4.105]:

[4.105] Turkish *hayir* ‘no’

ı̇pen dʒ’ o tʃobános, “**xaír.** dʒó ponáo ta...”
Said and the shepherd **no** not sell.PRES-1PS him
‘And the shepherd said, “**No**, I will not sell him.”’ (Phárasa dialect: 496)

Turkish *hich* was seen in example [4.73] in the sense of ‘not one’, and is used to mean ‘nothing’ in example [4.106]:

[4.106] Turkish *hiç* ‘nothing’

Andras **xítʃ** ré lai.
Man **nothing** NEG say.PRES-3PS
‘The man says **not a word.**’ (Silli dialect: 300)

The double negative found here is congruent with both Turkish and Greek.

Turkish *her* ‘each, every’ is found most often in these folktales modifying words relating to time, as can be seen in example [4.107]:

[4.107] Turkish *her* ‘each, every’

xér iméra esftʃánaʃi qavgá átʃi, “Daná ʃi sé ta ríʃes.”
every day do.PRES-3PP quarrel these calf your FUT it bind.2PS
“Gó sé ta rísu.”
I FUT it bind.1PS

‘**Every** day they used to have a dispute; “You shall tie up the calf.” “I shall tie it up.”’ (Silli dialect: 298)

The Turkish discourse-pragmatic borrowings found in Asia Minor Greek follow the patterns of Russian, Arabic, and Persian discourse-pragmatic borrowings in Tatar, and of Slavic borrowings found in Karaim. Despite the fact that Asia Minor Greek has been presented as a the parade example of a heavily influenced language, one located on the extreme end of the scale of borrowing, in many sentences found in the folktales cited here, discourse-pragmatic words are the only borrowings to be found, and their frequency and placement are remarkably similar to Russian discourse-pragmatic words in Tatar, suggesting that code-mixing was the original mechanism of borrowing. Finally, as in Karaim, also a heavily influenced language, many of the borrowings remain true to their dominant-language grammatical requirements, such as word order (*ichin, mi*) and vowel harmony (*da, mi*).

4.4.2 Other aspects of Turkish influence in Asia Minor Greek

Like Karaim in contact with Slavic, Asia Minor Greek was clearly in the third stage of lexical borrowing and structural borrowing in the early 20th century when it was documented by Dawkins; in the terminology of this model, speakers of Asia Minor Greek were in Generation D. In addition to prototypical content words such as nouns and verbs and discourse-pragmatic borrowings, we find other borrowings from Turkish as well. These additional borrowings include grammatical words without discourse-pragmatic functionality, Turkish-style reduplication as a means of intensification, and Turkish-style separation of action and manner.

Grammatical borrowings

In addition to the 47 discourse-pragmatic words borrowed from Turkish into Asia Minor Greek, three non-content words were borrowed that do not have metalinguistic function:

[4.108]

- *kanda* ‘where?’, in Greek as [gan] and [kandi]
- *karshi* ‘opposite’, in Greek as [xarʃú]
- *lazim* ‘necessary’

The one Turkish deontic found in Greek, *lazim* ‘necessary’, is a twice-borrowed word, originally borrowed into Turkish from Arabic. *Kanda* ‘where’ is the only Turkish question word borrowed into Greek, and is used in a purely spatial meaning. (Compare this to *kak* ‘how’, the only Russian question word code-mixed in Tatar, which is used only in its

evaluative meaning, and never to question manner.) *Kārshī*, a Turkish postposition meaning ‘opposite’, is used in Greek in its purely spatial meaning, as seen in example [4.109]:

[4.109] Turkish *karshī* ‘opposite’

op **xarśú** soruŋi perju tous ta **serája**.
 from **opposite** see.PRES-3PP son.GEN them the **palace.PL**
 ‘**Before** them they see their son’s **palaces**.’ (Silli dialect: 290)

Although *kanda*, *karshī* and *lazīm* are the only three grammatical words without discourse-pragmatic function found in the texts collected by Dawkins, based on the presence of other borrowing phenomena that place Asia Minor Greek in Stage Three of lexical and structural borrowing, it seems likely that other grammatical words were also borrowed from Turkish into Asia Minor Greek.

Intensification

In section 4.4.1 above, we saw several Turkish discourse-pragmatic words used explicitly for intensification, most notably *pek* ‘very’. However, Asia Minor Greek borrowed more than just Turkish discourse-pragmatic words to express intensification – it also borrowed Turkish-style reduplication. In Turkish, as in other Turkic languages, intensification of adjectives is often expressed through initial-syllable reduplication, where the final consonant of the copy syllable is unpredictable – and famous among phonologists for its unpredictability – but quite often [p]. For colors in particular, this is the preferred means of intensification in Turkish. An example of this intensifying Turkish initial-syllable reduplication is found in [4.110]:

[4.110] Initial-syllable reduplication to intensify *kırmızı* ‘red’

batırsen do so leró mésa, ke éne **qıp-qırmızı.**
dip.SIMPPST-3PS her in water inside and become.SIMPPST-3PS **VERY red**
‘She dipped her in the water and she became **all red.**’ (Araván dialect: 334)

This sentence is otherwise quite Greek in its morphosyntax: the verbs are not clause-final; there is the characteristically Greek use of doubled spatial relators before and after the noun (here *so léro mésa* lit. ‘in water inside’, or ‘into the water’; also seen in example [4.79] as *so miván apáno* lit. ‘in fruit tree on’, or ‘in the fruit tree’); native Greek *ke* ‘and’ is used for narrative structure; there is just one Turkish lexical item in the sentence. Therefore one might think that this reduplicated [qıp-qırmızı] was borrowed as a single unit meaning ‘very red,’ without sensitivity to the Turkish morphological process. However, several sentences later in the same narrative, we find this:

[4.111] Initial-syllable reduplication to intensify Greek *mávro* ‘black’

batırsen do so mávro so leró, ke épken do **mas-mávro.**
dip.SIMPPST-3PS her in black in water and make.SIMPPST-3PS her **VERY black**
‘She dipped her into the water, and made her **jet black.**’ (Araván dialect: 334)

Here we see that Greek *mávro* ‘black’ is intensified just as if it were a Turkish color term, by means of a reduplicated initial syllable rather than with a Greek or even Turkish free morpheme. There is a clear analogical base in Turkish for *mas-mávro*, shown in [4.112]:

[4.112] Analogical base for *mas-mávro*

Turkish:

mavi ‘blue’ : *mas-mavi* ‘very blue’

::

Asia Minor Greek:

mávro ‘black’ : X > *mas-mávro* ‘very black’

Clearly, despite the otherwise quite Greek nature of the sentences shown above, speakers of this dialect were sensitive both to the Turkish morphological process and to the fact that this reduplication is the preferred way to intensify colors, and expanded the use of initial-syllable reduplication to the Greek color term.

Initial-syllable adjectival reduplication is not the only kind of Turkish reduplication borrowed into Asia Minor Greek, as can be seen in example [4.113]:

[4.113] Turkish-style full reduplication

Ke kalá kalá **jashatse.**
and well well **live.AOR-3PS**
‘And she **lived** in great prosperity.’ (Ulaghátsh dialect: 378)

This sentence is the final sentence of a folktale, describing the happily-ever-after ending, and the speaker has used full reduplication of the Greek adjective to emphasize just how happy this ending is – note that this is not a Turkish adjective that can be interpreted as “staying Turkish” by being fully reduplicated for emphasis, but rather the absorption of the Turkic model of reduplication into Greek. Other examples of fully reduplicated Greek adjectives include: *pikrá pikrá* ‘very bitterly’ (Silli dialect: 286), *polí polí* ‘very’ (Silli dialect: 302), and *afíka afíka* ‘very gently’ (Phárasa dialect: 498). Just as initial-syllable reduplication turned out to not be limited to the stratum of Turkish borrowings, the use of full reduplication for

intensification does not seem to be either – the Turkish mechanism of reduplication for intensification seems to have been borrowed and generally applied.

Manner

In addition to Turkish-style reduplication, Asia Minor Greek speakers also used Turkish models and lexical items to describe manner. One such importation from Turkish was the use of (frozen) reduplicated gerunds or participles functioning as adverbs that describe the manner of a verb, usually a verb of motion, as seen in example [4.114]:

[4.114] Turkish-style adverb

Ekí	to	qutjí	yze yze	píjen	s	éna	mikró	xorjós.
there	the	box	floating floating	go-PST.3PS	to	a	small	village

“That box by **floating and floating** came to a little village.” (Delmesó dialect: 316)

Unlike the adjectival reduplication described in the previous section, which became a productive morphological process used for Greek words as well as Turkish, these reduplicated gerunds or participles were limited to the Turkish stratum of Asia Minor Greek, and were possibly borrowed as a single unit, as the only adverbial phrases of this type found in the folktales are with Turkish lexical items. A particularly common example of this Turkish adverbial form is found in [4.115]:

[4.115] Turkish-style adverb

Ge	to	peđí	estáθen	ɟaʃkin ɟaʃkin	so	dengizju
and	the	boy	stand.PST-3PS	wondering wondering	on	lake.GEN

so **kener.**
on **shore**

‘And the boy stood **wondering and wondering** on the **shore** of the **lake.**’ (Silli dialect: 308)

While this means of describing manner with frozen reduplicated gerunds and participles did not become productive in Greek, such that the reduplication process was applied to Greek verbs as well, what does seem to have become productive is the Turkish-style separation of manner and action, where the manner of action is described with one verb (or verbal form, as seen above) and the action itself is denoted with another. This separation of manner is distinctly Turkic (it is also found in Tatar¹²), and can be seen in example [4.116]:

[4.116] Turkish-style separation of manner and action

“Pulo mıla” dı, tşiyirda.
 sell.PRES-1PP apple say cry.pres-3PS
 ‘“I sell apples” she cries.’ (Sılata dialect: 442)

The modern standard Turkish equivalent is shown in [4.117]:

[4.117] Turkish equivalent

“Elma satıyor-u-m” diy-e baair-a
 Apple sell.PRES-1PP say-GER shout.PRES-3PP
 ‘“I sell apples” she cries.’ (Ayla Algar, personal communication)
 More literally, ‘Saying “I sell apples” she shouts.’

Both the use of the verb ‘to say’ after quoted speech and the separation of manner from action found in the Greek sentence can be seen as modeled on Turkish.

¹² It is not found in the Karaim used as data for section 4.2. However, it is not clear if this is the result of having so few example sentences to work with, or if the loss of this Turkic separation of action and manner is another form of Slavic influence found in Karaim.

4.5 Conclusions

We have seen that sociolinguistic context and discourse pragmatics can explain both the motivations for and mechanisms of certain kinds of lexical and structural borrowing. In language contact situations with similar asymmetries of power and functional domain, as well as the requisite level of bilingualism, words with the same discourse-pragmatic functions (and when there is a common dominant language, the very same words) are borrowed again and again. Appendix F of this dissertation gives a side-by-side comparison of all of the discourse-pragmatic borrowings discussed in both Chapter Three and the above case studies. Figures 17 and 18 below graphically represent two instances of this repeated borrowing and code-mixing. Figure 17 shows the repeated borrowing and code-mixing of words meaning ‘and’, and Figure 18 shows the repeated borrowing and code-mixing of words meaning ‘of course’. In these figures, vertical arrows are diachronic and represent borrowing, while horizontal arrows are synchronic and represent code-mixing.

Figure 17. Repeated borrowing of ‘and’

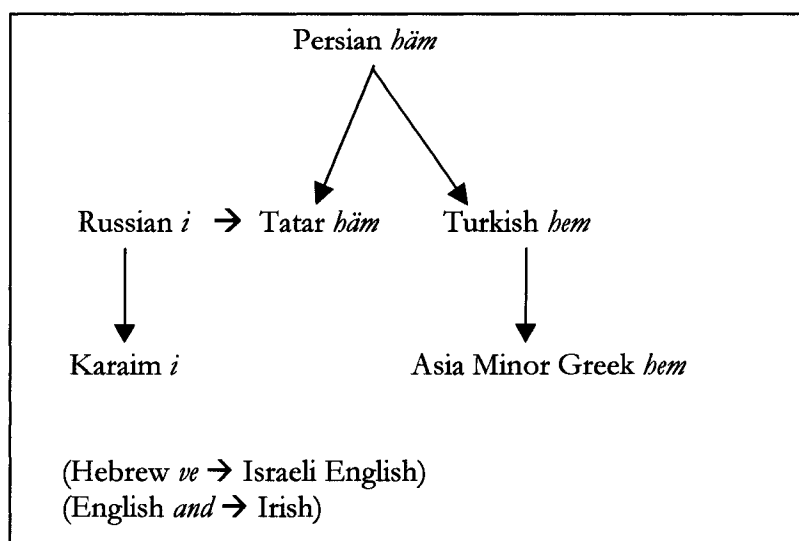
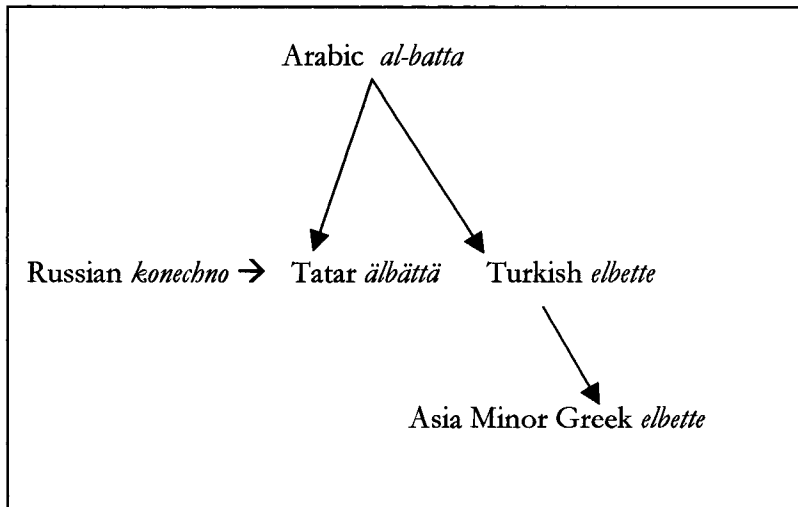


Figure 18. Repeated borrowing of 'of course'



Many code-mixed and borrowed discourse-pragmatic words seem to retain their dominant-language grammatical rules and requirements, such that their use is associated with a composite morphosyntax, and the structure of the minority language begins to increasingly resemble the structure of the dominant language:

- Russian subordinators in Tatar introduce Russian-style post-head subordinate clauses, and Persian subordinators in Tatar introduce Persian-style post-head subordinate clauses, rather than native Turkic embedded pre-head clauses. Conversely, Turkish subordinators in Greek govern Turkish-style pre-head embedded clauses, unless they are from Persian, in which case they are clause-initial and post-head (having retained their syntactic features through two instances of borrowing). The borrowed Persian subordinators *ägär* 'if' and *gärchä* 'although' mean that conditionality and concession are marked sentence-initially in Tatar, as they are in Persian, and not merely clause-finally in the verb, which is the native construction.

- The introduction of dominant-language coordinators can mean a change in the manner of coordination. Arabic *wä* ‘and’ and Persian *häm* ‘and’ are used in constructions in Tatar that have both different word order and syntactic relations than constructions using the native postposition *belän* or enclitic *DA* for coordination. Use of Russian *tozbe* ‘also’ in Tatar is inducing a move away from the doubled-enclitic *DA* – it is neither a postposed enclitic, nor used in each of the phrases that it is coordinating. Meanwhile, the borrowing of Turkish *da* into Asia Minor Greek introduced a coordinative clitic that is postposed after the verb that it is coordinating, in contrast to native Greek preposed coordinating conjunctions.
- The Russian interrogative particle *chto li* is always in the sentence-final word slot in Tatar, which is sometimes but not always congruent with the placement of the Tatar interrogative particle *mI*. The Turkish interrogative particle *mI* when used in Greek retains its word order requirements, and thus follows the word being questioned.
- Grammaticalized Russian prepositions borrowed as discourse-pragmatic words in both Tatar and Karaim retain their preposed position, while the native equivalents are postposed; e.g., *okolo* ‘around’, which is found in both Tatar and Karaim, causes a slight alteration the word order of both languages by remaining pre-head.
- The Russian particle *chut* ‘used in Tatar for ‘almost’ entirely changes the realization of the verb it modifies: it retains its requirement for a verb in the negative (which is not found in Tatar), and is used in lieu of an auxiliary verb rather than a particle.
- Russian *mozhet* ‘might’, borrowed into Tatar as an epistemic, causes a slight change in word order: it is placed adjacent to the word that is having its probability assessed, while native Tatar (actually < Arabic) epistemics are usually clause- or sentence-initial.

- Finally, the Turkish enclitics *da* and *mi* retained their adherence to vowel harmony when used in Asia Minor Greek, perhaps one of the mechanisms through which vowel harmony (found to some extent in all of the dialects of Asia Minor Greek) was introduced.

An examination of the lexical and structural borrowing phenomena found in Karaim, Tatar, and Asia Minor Greek suggests that the process of borrowing from dominant languages into minority languages is in some ways the temporal reverse of grammaticalization: discourse-pragmatic words are first borrowed *only* in their grammaticalized form, and only later in their root meanings, and grammatical words or morphemes without discourse-pragmatic function are only borrowed in Stage Three, after there has been a stage of discourse-pragmatic borrowing. Within the general rubric of grammaticalization, Traugott (1989) identifies three tendencies of semantic change: (1) Meanings that are based in an external described situation become meanings based in an internal situation, be it evaluative, perceptual, or cognitive; (2) Meanings based in an internal or external described situation become meanings based in a textual and metalinguistic situation; (3) Meanings tend to become increasingly based in the speaker's subjective belief state or attitude towards a proposition (this includes the development of epistemics from deontics and adverbs of stance from adverbs of manner). In general, words that are found at the endpoint of these three paths of semantic change are borrowed in Stage Two of the

model presented here, while words that have not undergone grammaticalization are borrowed in Stage Three.¹³ Table 5 summarizes these different stages of borrowing:

Table 5. “Reverse grammaticalization” in minority-language borrowing

Borrowed in Stage Two	Borrowed only in Stage Three
Epistemics <ul style="list-style-type: none"> • Russian <i>mozhet</i> ‘might’ in Tatar 	Deontics <ul style="list-style-type: none"> • Turkish <i>lazım</i> ‘necessary’ in A. M. Greek
Adverbs of stance <ul style="list-style-type: none"> • Russian <i>absolutno</i> ‘absolutely’ in Tatar • Arabic <i>motlakan</i> ‘absolutely’ in Tatar 	Adverbs of manner <ul style="list-style-type: none"> • Arabic <i>mötawazigan</i> ‘humbly’ in Tatar • Turkish <i>üze üze</i> ‘floating’ in A. M. Greek
Grammaticalized spatial relators used as hedges, evaluators of degree, etc. <ul style="list-style-type: none"> • Russian <i>okolo</i> ‘approximately, around’ in Tatar and Karaim • Arabic <i>kadär</i> ‘as much as’ in Tatar 	Spatial relators in their root meanings, at times the same words that have been borrowed in Stage Two in grammaticalized meanings <ul style="list-style-type: none"> • Polish <i>kolo</i> ‘by, near’ in Karaim • Arabic <i>kadär</i> ‘up to’ in Tatar
Question words with evaluative function <ul style="list-style-type: none"> • Russian <i>kak</i> ‘how’ in Tatar 	Question words without evaluative function <ul style="list-style-type: none"> • Turkish <i>kanda</i> ‘where’ in A. M. Greek
Coordinators used for narrative structure only <ul style="list-style-type: none"> • Turkish <i>hem</i> ‘and’ in A. M. Greek • Russian <i>i</i> ‘and’ in Tatar and Karaim 	Coordinators used as general grammatical coordinating conjunctions <ul style="list-style-type: none"> • Persian <i>häm</i> ‘and’ in Tatar

This “reverse grammaticalization” in borrowing is only *temporal*, and describes the necessary ordering of borrowings – I am by no means suggesting by this term that words are being de-grammaticalized to their root meanings and functions. It is not that Persian *häm* ‘and’ was borrowed into Tatar as a discourse marker of narrative structure and then in Tatar underwent “degrammaticalization” so that it also started to function as a general coordinating conjunction. Nor did grammaticalized Arabic *kadär*, borrowed as an evaluator of degree meaning ‘as much as’, then undergo a reverse semantic change in Tatar and begin functioning as a spatial relator meaning ‘up to’. Rather, these words were borrowed twice: in Stage Two in their grammaticalized meanings and functions, and in Stage Three in their root meanings and functions. It is not clear at this time if this change in borrowing patterns is due

¹³ Note that words on Traugott’s “propositional” level of language can have pragmatic functionality,

to some sort of change in the contact situation, for example, further contraction of the population of speakers or of functional domains or a further change in the asymmetry of power and prestige, or if Stage Three borrowings reflect a level of structural influence that occurs when a sufficiently large percentage of the population of minority-language speakers is bilingual and highly competent in the dominant language for a sufficiently long amount of time.

With all this in mind, let us revisit the model of lexical and structural borrowing laid out in the introduction of this chapter, and integrate the findings from the language contact situations examined here.

Stage One

- Does not require high level of bilingualism (neither number of speakers nor speaker competence)
- Only content words/denotative language borrowed (mostly nouns and verbs)
- Non-basic vocabulary borrowed first, then basic vocabulary borrowed
- No other borrowing phenomena

Stage Two

- Requires a reasonable level of bilingualism in the populace and competence in the dominant language for the minority speakers who are doing the borrowing

and often do not follow the above paths of semantic change and grammaticalization; these words, such as 'yes' and 'no', are borrowed in Stage Two.

- Code-mixing and eventually borrowing of discourse pragmatic words to structure and comment on narrative; perform social gestures; express speaker evaluation of and stance toward the discourse
- Polysemous words borrowed only in grammaticalized meanings with discourse-pragmatic functionality: includes borrowing of epistemics, adverbs of stance, grammaticalized spatial relators that are used as hedges and other evaluatives, question words with evaluative function, coordinators used for narrative structure. Only temporal and spatial deictics with discursive function will be borrowed.
- Borrowing of free morphemes, including enclitics
- Retention of dominant-language discourse-pragmatic words' native grammatical requirements leads to a composite morphosyntax that is more like dominant-language grammatical structure, including but not limited to:
 - changes in subordination patterns, where languages with embedded pre-head subordinate clauses acquire extrapolated post-head subordinate clauses, and vice-versa
 - syntactic changes in coordination
 - change in syntactic expression of interrogatives, including word order of interrogative particles
 - introduction of preposed modifiers in a postposing language, and vice-versa
 - change in expressions of epistemics, including alterations in word order and the use of epistemic verbs rather than adverbs
 - changes in the syntactic expression of conditional and concessive clauses
 - introduction of vowel harmony through enclitics

Stage Three

- Level of bilingualism at least equal to but most likely greater than level of bilingualism found in Stage Two
- Borrowing of grammatical words without discourse-pragmatic functionality: includes deontics, adverbs of manner (including gerundial forms used as adverbs of manner), spatial relators in their root meanings used as general prepositions or postpositions (in accordance with dominant-language word order), question words without evaluative function, coordinators used as general grammatical coordinating conjunctions
- Borrowing of personal deictics, and spatial or temporal deictics without metalinguistic function
- Borrowing of bound morphemes
- Further structural influence from the dominant language, including such phenomena as:
 - use of dominant-language mechanisms of intensification such as initial-syllable and full reduplication in both native and borrowed strata of the lexicon, e.g., Asia Minor Greek borrowing of Turkish initial-syllable and full reduplication
 - copying of dominant-language expressions of manner, e.g., Asia Minor Greek borrowing of Turkish separation of manner and action
 - use of the inflectional morphology required by the parallel construction in the dominant-language, e.g., Karaim use of the genitive suffixes *-nIn* (native) and *-a* (Russian) in constructions with Russian prepositions that govern the genitive case
 - creation of cases parallel to dominant-language cases, e.g., innovated use of the Karaim enclitic *bə* as a case marker parallel to the Slavic instrumental

- calquing of dominant-language coordinating constructions, e.g., innovated use of the Karaim enclitic *da* to parallel Slavic coordinating conjunctions
- expansion of polysemy patterns parallel to the dominant language equivalents, both semantic and syntactic, e.g., innovation of Karaim *kayda* ‘where’ as a relativizer, parallel to Russian *gde* ‘where’
- change in overall word order, e.g. Karaim SOV > SVO

This model incorporates discourse theory to offer plausible motivation for lexical and structural borrowing, and shows that universal tendencies of grammaticalization can be used to explain some of the hierarchies of borrowing in language contact situations with an asymmetry of power, prestige, and breadth of functional domain. In Chapter Five, I will propose several potential motivations for the code-mixing and borrowing of discourse-pragmatic words, and suggest ways in which these hypotheses can be tested with future research.

Chapter Five

Conclusions

The analysis of the Tatar spoken and written in post-Soviet Tatarstan presented in this dissertation demonstrates that it is necessary to look outside of language in order to understand both linguistic structure and structural innovations. The linguistic performance of young Tatars can only be understood when contextualized at both the micro-social and macro-social levels, and shows that one must analyze language as a practice in order to accurately describe language as a system.

In Chapter One, I presented the broad socio-political context of post-Soviet Tatarstan and then located individual linguistic performance within this larger context. A brief historical overview of the Tatar nation and Tatar-Russian contact showed that although for many Tatars the fall of the Kazan Khanate in 1552 is considered to be the most significant moment in Tatar history, the Tatar nation was relatively powerful and influential – with cultural, religious, educational, and socio-economic ties to the wider Islamic and Turkic worlds – up until the beginning of the Soviet era. The period from the mid-19th century until the 1917 revolution in particular is considered to be the golden era of Tatar culture, characterized by a liberal and modernizing reform movement, highly regarded writers and philosophers, and the adoption of Kazan Tatar as a literary language by neighboring Turkic peoples. It is the Soviet period, with its purges of the Tatar political, cultural, and literary elite, the systematic destruction of Tatar educational and religious institutions and publishing houses, and the isolation of Russian Tatars from their Turkic brethren in Central Asia and Turkey that most significantly undermined the prestige, influence, and stability of the Tatar nation. The covert Soviet policy of “Russification” – which included a forced move to a modified Cyrillic alphabet, language engineering where

Russian was the only viable source of linguistic “enrichment,” and the limitation of Tatar’s functional domains – appears to have played a significant role in changing the situation of Tatar in Tatarstan from one of stable bilingualism to one where Tatar is a “contracting” language, losing both speakers and functional domains. Much of this Soviet legacy remains in post-Soviet Tatarstan: Tatar is to this day particularized in function and stigmatized by many, although it does retain “covert prestige” among much of the Tatar intellectual urban elite, and the language shift of the Tatar populace appears to be continuing apace and is thus classifiable as “gradual language death.” In an effort to combat this stigmatization and language shift, the ethnically Tatar government of post-Soviet Tatarstan has been engaged in the promotion of the Tatar nation as part of their sovereignty project, and although the official national ideology is one of civic multiculturalism, government resources are used to promote Tatars over other ethnic groups in the republic. Language legislation and policy have been an important part of this systematic promotion of Tatars: the declaration of sovereignty in 1990 raised Tatar to the status of an official language, and the government has promoted the mandatory teaching of Tatar in schools, an increase in the availability of Tatar-language education, and the study of Tatar by members of other ethnicities. However, particularly among the Tatar intelligentsia, there is a sense that language equality is *de jure* rather than *de facto*, and a survey of Tatar’s functional domains shows that this is indeed the case. There is an asymmetry of use and usefulness for Russian and Tatar; for example, Russian is the dominant language of the media (newspapers, magazines, television, and radio), education at all levels, and the public sphere in general.

The individual linguistic performance of Tatar urban bilinguals is set within this context of Russian dominance: the sociolinguistic conventions of present-day Tatarstan require accommodation to Russian in public domains, and this accommodation is part of the style shifting of these urban bilinguals. Violation of the rules of linguistic etiquette, most notably

using Tatar in situations that call for Russian, is used by some Tatars as a symbolic gesture: the refusal to accommodate to either Russian public space or monolingual Russian speakers demonstrates a stance that prioritizes ideology over communication, and metonymically represents the refusal to accommodate to Russian and Russians in domains outside of language. Bilingual Tatar speakers have a range of styles that can be organized according to level of language mixing, with pure Tatar on one end, pure Russian on the other end, and a variety of “mixed” styles in between; these styles are used for the audiences and situations for which they are appropriate. The theory of audience design can be used both to analyze this style shifting and to systematically account for the effect of the fieldworker on both observed and recorded speech: both the participant role of the fieldworker and the social role that is assigned to her by conversation participants can be used to explain their style shifting and style choice. Tatar styles vary in their level of “verbal hygiene,” highest in the most “performed” style, which I have labeled “Tatar on-stage” style, most saliently characterized by its “de-Russification.” “Performed” styles, which are understudied, are in fact remarkably useful for sociolinguistic and linguistic anthropological inquiry, and the dichotomy between “natural/unnatural” speech is a false one, as all speech is in some way “performative.” In addition, an analysis of Tatar style shifting demonstrates the necessity of multi-stylistic data: the different styles of Tatar bilinguals demonstrate different levels of Russian influence and content, and if a description of Tatar linguistic structure were based on just one of these styles, it would be erroneous.

Individual style shifting is set against the backdrop of post-Soviet Tatar language ideologies. Chapter Two shows how language ideologies are the mediating link between Tatar’s socio-cultural context and its linguistic structure. Language ideologies are never simply about language, and the language ideology that I call the “Tatar discourse of purity” can be seen as a counter-hegemonic discourse that expresses resistance to Russian domination. This discourse of

purity, as is common in post-colonial and post-imperial transitions, is part of the attempt to “revalorize” a subordinate group, and language is used as a symbolic resource. The Tatar discourse of purity is a significant part of the construction of the post-Soviet Tatar identity, and is based on both a logic of “oppositional identity” and “selective targeting” – Tatar identity is defined in opposition to Russian, and Russian is the selective target of purification. “Pure” Tatar is implicitly defined as Tatar without Russian influence or interference, and is implicitly opposed to the “impure” Tatar that is ubiquitous in Kazan – Tatar with phonetic interference, spelling mistakes, or errors that occur in translation from Russian. Code-switching with Russian is also perceived as “impure” Tatar. While Russian influence is rejected, influences from languages that are not perceived as threats can be interpreted as congruent with the post-Soviet Tatar identity. Language is a marker of group identity and often part of a feature cluster that includes descent, religion, and culture: however, at least in post-Soviet Tatarstan, language in particular is seen as a metonymic representative of the nation, where the impurity and decline of the Tatar language are seen as representative of the impurity and decline of the Tatar nation as a whole. The expression of the Tatar discourse of purity on an individual level is found in the de-Russified Tatar on-stage style; on the communal level, it is found in purification movements that are common in post-colonial contexts, most notably orthographic and lexical reform. The recently legislated shift from a modified Cyrillic alphabet to a Latin-based alphabet can be read via metaphor and metonymy as a symbolic gesture of cultural and political reorientation, away from Russian and Russia and towards both Turkey and the West. Other historical Tatar orthographies can also be linked via metaphor and metonymy to an interest in the cultures, politics and social realities associated with them: pan-Turkic runic writing is associated with both the antiquity of the Tatar nation and with pan-Turkic separatist leanings, while Arabic orthography is used both to express religious orientation and to allude to the “golden era” of pre-Revolutionary Tatar

culture. The purification of the lexicon can also be read symbolically: many Russian words brought into the Tatar language during the Soviet period are being discarded, and their replacements are usually not native Tatar words, but rather the Arabic loanwords that were removed from Tatar by Soviet-era language engineering. Once again, Russian is the selective target, while Arabic influence, part of the pre-Soviet Tatar identity, is seen as congruent with the post-Soviet Tatar identity as well. Post-Soviet Tatar lexical reform also shows which elements of Russian influence are most salient to speakers and thus subject to purification on the individual level: the words in question are all content words and the vast majority are nouns. An examination of 20th century post-imperial Turkish lexical and orthographic purification movements shows that they too were based on revalorization, oppositional identity and selective targeting: here, Arabic orthography and lexical items were the targets of purification. This comparison with the Turkish case makes particularly clear the ways in which socio-cultural and political structure can affect linguistic structure: language engineering in both Turkey and Tatarstan has rendered texts written just one hundred years ago nearly unintelligible to modern speakers, and has reduced the mutual intelligibility of the two languages.

In Chapter Three, I returned to the Tatar discourse of purity, integral to the understanding of the linguistic performance of individual Tatars, who have “pure” (= “de-Russified”) Tatar at one extreme of a cline of language mixing. I have labeled this style “Tatar on-stage” style, as it is a “performed” style integral to the construction of a cultural and ethnic Tatar identity. This style, where speakers exhibit the highest level of verbal hygiene, is found when audience and setting require a prominent presentation of Tatar identity; for example, public speaking at Tatar cultural events, all speech on Tatar radio and television, regardless of register, and conversations with investigators of Tatar language and culture. When the audience and setting require Tatar to be the language of conversation, but do not require such a high level

of verbal hygiene, young Tatars will use what I call “Tatar-preferred style”: while they are still filtering out Russian elements, these Russian elements are limited to content words, while Russian discourse-pragmatic words on the metalinguistic level of language are allowed to “sneak through.” Neither speakers nor listeners appear to be aware of these Russian words that are organizing and commenting on otherwise purely Tatar discourse, and they can even appear (to a much lesser extent) in the highly performed Tatar on-stage style. This code-mixing is particularly interesting because it is found in the speech of even the most culturally and politically active young Tatars in Tatarstan, speakers who are either Tatar-dominant or balanced bilinguals and who use Tatar in a variety of functional domains, including in professional and educational spheres. Evidence of this code-mixing was found in the speech and writing of Tatar Social Club members who have demonstrated that language ideologies can be more important to them than communication, and who have explicitly stated that they dislike speaking Russian and use it as infrequently as possible. Many of these Russian discourse-pragmatic words retain their Russian grammatical requirements when they are code-mixed into otherwise Tatar discourse; for example, in lieu of embedded pre-head clauses subordinated with non-finite verb forms we find post-head clauses introduced by subordinating conjunctions, and prepositions retain their prepositional nature (Tatar is postpositional). Tatar-preferred style can thus be analyzed as a “composite” that is mostly Tatar yet includes not only Russian lexical items, but also Russian semantic and syntactic patterns. This form of language mixing and the lexical and structural alterations that go along with it indicate a pathway of language attrition and language influence that has not been described in other studies of language contact.

Chapter Four shows that this Tatar pattern of language mixing is not an isolated case, and that using discourse-pragmatics as an organizing principle can both explain and give motivation for lexical and structural borrowing in other sociolinguistically similar contact

situations. Previous work on language contact has aptly typologized hierarchies of lexical and structural borrowing without giving real motivation for these hierarchies. A new three-stage model of lexical and structural borrowing, which is based not on the usual dichotomy of system/content but rather on the trichotomy of system/content/discourse-pragmatic, shows that while the borrowing of content words (Stage One) minimally alters a language, it is the borrowing of discourse-pragmatic words that retain their grammatical requirements which triggers contact-induced structural innovations (Stage Two). Additionally, cross-linguistic evidence demonstrates that discourse-pragmatic words with the same meanings and functions are borrowed again and again, and when separate contact situations have a dominant language in common, it is the very same words that are borrowed. For example, Persian *ki* ‘that’ can be found as a loanword in Tatar, Karaim, and Turkish, and was borrowed from Turkish into Asia Minor Greek, while Russian *chto* ‘that’ is code-mixed into Tatar. Coordinating conjunctions meaning ‘and’ are either code-mixed words or conventional borrowings in all of the languages examined here: Persian *häm* was borrowed into Tatar and Turkish, and borrowed again from Turkish into Asia Minor Greek; Tatar *wä* is a loanword from Arabic; and young Tatars currently code-mix Russian *i*, which is a conventional borrowing in Karaim. In addition, Hebrew *ve* is found code-mixed in Israeli English, and English *and* is found code-mixed in Irish discourse. Structural innovations induced by the code-mixing and borrowing of discourse-pragmatic words cause the structure of minority language to become more like that of the dominant language with which it is in contact, and include alterations in subordination, coordination, interrogatives, word order, verbal constructions, case governance, semantics, and polysemy patterns. Only after a language has gone through Stage Two borrowing of discourse-pragmatic words will it begin to borrow system words and morphemes (Stage Three). Discourse-pragmatic words that are polysemous will be borrowed only in their grammaticalized form in Stage Two, and in Stage Three may be

borrowed in their root meanings and functions. Grammaticalized borrowings in Stage Two include epistemics, adverbs of stance, spatial relators used as hedges, question words with evaluative function, and coordinators used for narrative structure. In Stage Three, borrowings include deontics, adverbs of manner, spatial relators used in their root meanings, question words without evaluative function, and coordinators used as general grammatical coordinators. Both discourse theory and theories of grammaticalization, usually applied to the analysis of monolingual data, are here used as organizing principles to distinguish between borrowings found in different stages of language contact and influence.

While introducing the category of discourse-pragmatics to the established categories of system and content words allows for a better understanding of the stages, hierarchies, and types of lexical and structural borrowing, an important question remains: just why are dominant-language discourse-pragmatic words borrowed into minority languages that have previously borrowed only content words? The mechanism of their borrowing appears to be code-mixing: the frequency and placement of the Russian discourse-pragmatic words code-mixed in Tatar parallel the frequency and placement of discourse-pragmatic words that are conventional borrowings, including Arabic and Persian borrowings in Tatar, Russian and Polish borrowings in Karaim, and Turkish borrowings in Asia Minor Greek. The question then should perhaps shift to code-mixing – why are dominant-language discourse-pragmatic words code-mixed into minority-language speech and writing where the only other dominant-language words are content words that have become conventional borrowings? Why does Tatar-preferred style seem to demonstrate that young Tatar bilinguals use Russian to organize and evaluate their Tatar discourse, even when they are Tatar-dominant, explicitly state that they prefer to not speak in Russian, and indicate through their linguistic performance that “pure” Tatar is an explicit and implicit ideal? There are several potential explanations for this behavior: one is cognitive, and

based on the peripherality of discourse-pragmatics words to awareness; a second highlights the iconicity of discourse-pragmatic words; and a third incorporates social factors into the use of dominant-language discourse-pragmatic words.

Discourse-pragmatic words, particularly those words that structure discourse, often come in peripheral word slots, and when they are not utterance- or sentence-initial they are often clause initial. It appears that discourse-pragmatic words are peripheral to awareness as well, and less salient than content-word, or denotative-language, nonce borrowings. Errington, writing on the syncretic composite language composed of Indonesian and Javanese known locally as “language salad,” contends that content-word borrowings are “conspicuous” to speakers and listeners, while discourse particles are not (1998: 10, 100ff):

“Extensive repertoires of discourse particles serve Javanese Indonesians as a means for marking feelings about and stances towards conversational topics, contexts, and participants. Their non-referential, crucially situated significances appear to make them peripheral for speakers’ awarenesses relative not just to their encoded linguistic functions...but also with respect to their various provenances as well” (ibid. 10).

Errington believes that these discourse particles function as what Gumperz calls “contextualization cues,” which are “[h]abitually used and perceived but rarely consciously noted and almost never talked about directly” (Gumperz 1982: 131, cited in Errington 1998: 100). This was seen in Chapter Three; young Tatars are apparently unaware that their “pure” Tatar linguistic performance was actually replete with Russian discourse pragmatic words, and this lack of salience seems to be repeated among the Karaim.¹ Articles in the post-Soviet Tatar press that are concerned with linguistic purity and the use of either Russian or Arabic borrowings refer *only* to denotative-level content words, and usually only to nouns – discourse-pragmatic words do

¹ When discussing linguistic purism among the Karaim, Csato’s sole example of a borrowing salient to speakers is a Russian noun, *stol* ‘table’ (1999a: 85).

not appear to be salient to Tatar speakers concerned with lexical reform. Nor do they appear to have been salient to Turkish language reformers in the first half of the 20th century: despite the fact that hundreds of Arabic and Persian borrowings were removed during the transition from Ottoman Turkish to Modern Turkish, the vast majority of Arabic and Persian discourse-pragmatic borrowings remained, and those Arabic and Persian discourse-pragmatic words that were borrowed into Asia Minor Greek via Turkish over a century ago are all in standard usage in Turkish today. Minority-language speakers who are communicatively competent in two codes, even those with an explicit discourse of purity, simply appear to not notice when they import the metalinguistic code of the dominant language into the denotative code of the minority language. The mixing of denotative codes is salient to speakers, but the mixing of a metalinguistic code and a denotative code apparently is not.²

Iconicity has also been suggested as an explanation for the code-mixing of discourse-pragmatic words into minority-language speech. Maschler's analysis of the Hebrew-English code-mixing briefly described in Chapter Four relies upon the separation of the denotative and metalinguistic levels of language; she contends that code-mixing is iconic and allows one language to serve as 'commentary' on the other (1994: 325). Maschler found that Hebrew code-mixing was "frequently employed in utterances that negotiate a boundary between the immediately preceding part of the interaction and the immediately following part" (ibid. 332). The metalinguistic code-mixed words thus segment the discourse by acting as boundaries between what Maschler calls "verbal activities"; for example, Hebrew discourse-pragmatic words are used to separate interruptions and asides from the main discourse in English. The discourse-

² It is currently unclear why this apparent difference in salience should exist, and this is a question that invites cognitive linguistic research into the processing of discourse organization.

segmentation function of these words is therefore enhanced by their being in a different language: the words themselves create discourse boundaries, and the switch to another language emphasizes the fact that a boundary has been created. Discourse-pragmatic words used to express contrast can also be interpreted as iconic. For example, Cotter notes that the English word *but* code-mixed in Irish discourse is contrastive in two ways: it is semantically contrastive, and it is contrasted with the Irish discourse that surrounds it by means of being in English (1996: 187).

However, code-mixed discourse-pragmatic words do more than just segment discourse and express contrast: they structure discourse in other ways such as subordination and coordination; they organize event structure and are used for deixis and as directives; they perform social functions such as greeting and leave taking; and they express both speaker stance toward and evaluation of the denotative discourse at hand. Therefore iconicity, while a plausible motivation, cannot be the only explanation for their use. Additionally, while the cognitive explanation appears to be extremely plausible – when speakers are engaged in verbal hygiene and filtering their speech, discourse-pragmatic words are salient and thus removed only at the very highest level of verbal hygiene and purity requirements – this explanation too seems insufficient in some ways. Perhaps cognitive reasons alone can explain cases of interference where speakers import the discourse organization from the language in which they feel most comfortable into a language in which they have a lower level of competence. However, the question remains: why are Tatar-dominant and balanced bilingual speakers – who actively demonstrate that they prefer speaking Tatar – importing Russian discourse organization and evaluation into Tatar? A third hypothesis, which requires a significant amount of additional research in order to be tested, suggests that in addition to their iconicity and lack of salience for language filtering, dominant-language discourse-pragmatic words may be used because of the additional discursive and

pragmatic force that comes from their association with a language of prestige and power. This hypothesis is based on several considerations. The first comes from work by Jane Hill and Kenneth Hill on *Mexicano*, a “syncretic code” based on Nahuatl and Spanish with both a discourse of purity and varying levels of language mixing. They suggest that hispanization “evokes the power and prestige of Spanish-speaking society, and lends dignity and force to Nahuatl utterances” (Hill and Hill 1980a: 122). In addition, they found a correlation between the expression of stance and the use of the dominant language; the more evaluative a narrative was, the more Spanish and hispanisms were found in that narrative (1980a, 1980b).

“Analysis of narratives shows that the highest rates of hispanization are consistently found in what Labov (1972) calls ‘evaluation’: the aspects of narrative discourse through which the speaker shows why his narration is worthy of the attention of the hearer... Generally, in types of discourse hispanization and switching is dominant in evaluative and explanatory material and apparently lends dignity, force, and convincingness to the utterance by its presence” (Hill and Hill 1980b).

Hill and Hill found that speakers of *Mexicano* generally used a level of hispanization commensurate with their prestige level, and suggest that low-prestige men and women “monitor down” their rate of hispanisms (1980a: 128); note that this type of verbal hygiene is the reverse of what we see among young Tatar speakers.

Another phenomenon suggesting that dominant-language discourse-pragmatic words are being code-mixed for additional power and emphasis is their use in double-marking. In Asia Minor Greek, we saw the phrase [pék polí kaló] ‘very very good’ (example 4.99), where *polí* is the native Greek intensifier ‘very’ and Turkish *pek* ‘very’ is used for additional intensification. Persian *ägär* ‘if’ is used to double-mark conditionality in Tatar, introducing a clause where the verb is already in the conditional, and Persian *gärchä* ‘although’ is used to double-mark concession, introducing a clause where the verb is in the concessive form created by the combination of the

conditional affix *-cA-* and the enclitic *DA* ‘also’. Young Tatars will periodically double-mark their evaluation of degree by using the phrase *daxbe kbätta* ‘even’, where both Russian *daxbe* ‘even’ and Tatar *kbätta* ‘even’ are used, and also often double-mark their greetings and leave-takings, saying “*Sau bul. Poka.*” ‘Bye. Bye.’ This double-marking suggests a phenomenon similar to monolingual cycles of emphatic markers, where one intensifier becomes semantically bleached and thus not sufficiently emphatic, and a new as-yet-unbleached intensifier is chosen (as can be seen in the set of English words used as intensifiers, e.g., *awfully*, *terribly*, *horribly*, *bella* (dialectal) and *wicked* (dialectal)). Semantic bleaching, where a word loses its semantic power, is part of the process of grammaticalization. Is it possible that there is an analogous “bleaching” of power in low-prestige and contracting languages, here a bleaching of pragmatic and discursive force rather than semantic? In monolingual cases of semantic bleaching of intensifiers, speakers use native resources to generate new intensifiers with sufficient power. Perhaps, and this is at present mere speculation, for speakers of contracting minority languages the native resources of the language are no longer sufficiently powerful due to the loss of what Bourdieu calls “symbolic capital,” and they turn to the resources of the dominant language instead, as it has increased its symbolic capital at the expense of the contracting language. Many of the dominant-language discourse-pragmatic words found in the examples given in Chapters Three and Four are words that would be emphasized prosodically in monolingual English discourse: for example, in Asia Minor Greek we find: “**Would that** I had not killed my daughter” (example 4.90); “**Come**, bring your youngest daughter” (example 4.91); “**No**, I will not sell him” (example 4.105). Perhaps Greek speakers, instead of using native means of emphasis and stress, sometimes used discourse-pragmatic terms from Turkish to lend their statements sufficient power and force.

A fourth and final hypothesis is that the code-mixing of dominant-language discourse-pragmatic words is one way of creating a “syncretic code” that constructs a desired identity

without explicitly violating a prevalent discourse of purity. The term “syncretism” was originally used by Kurylowicz to describe linguistic structures that suppress oppositions, and was first adapted to the study of bilingualism by Jane Hill and Kenneth Hill in their work on Mexicano. Both the Mexicano analyzed by Hill and Hill (1980a, 1980b, 1986) and the Indonesian “language salad” analyzed by Errington (1998) are described by them as syncretic codes, defined as a “‘mixed’ usage which suppresses the social relevance of oppositions between systems and in which the provenances of talk’s elements...are interactionally muted” (Errington 1998: 9). Cotter interprets her Irish speakers’ insertion of English discourse markers into otherwise Irish radio performance as “foregrounding the bilingual foundation upon which interlocutors build their discourse” (1996: 188) and “as a linguistic marker of bilingual identity” (ibid. 195). Perhaps Tatar-preferred style is used in private conversations because the entirely de-Russified Tatar on-stage style would be inappropriate due to its associations with the “performance” of Tatar identity, which is appropriate for public speaking and conversations with outsiders, but not for conversations among friends, where it might be perceived as artificial, overly formal, or distancing. If de-Russified Tatar is construed as a “performed” style that is inappropriate and code-switching is also inappropriate for the audience or situation, perhaps the code-mixing of Tatar-preferred style can be interpreted as due, in part, to audience design. Because there is no code-mixing of content words, this style does not saliently violate the explicit discourse of purity found in the Tatar community, but perhaps the code-mixing of discourse-pragmatic words references the bilingual foundation of the speakers and acknowledges that they are *choosing* to speak Tatar although they are fully capable of navigating their way through a Russian-dominant world.

These findings and hypotheses clearly invite additional research and analysis. Particularly desirable is a broad typological study of code-mixing and lexical borrowing in a variety of

contact situations. Is the three-stage model of hierarchical borrowing described here true for all types of contact situations? What role does discourse-pragmatic functionality play in borrowings from substrates, or between languages with a much less severe asymmetry of power and function? If increased discursive and pragmatic force via association with the dominant language is indeed a motivation for code-mixing of discourse-pragmatic words, this predicts that different patterns of code-mixing and borrowing will be found in different sociolinguistic contexts. We would then expect that in contact situations with a less marked asymmetry of power and prestige, we would find more code-mixing and borrowing of discourse-structuring words than of evaluative words, as the “iconicity” of boundary-creation merely reflects that the words are in a different language, regardless of the relative prestige of the languages. If code-mixing and borrowing of discourse-pragmatic words is essentially the same for typologically different contact situations, then the “additional force” hypothesis can be discounted. Additional research on the Tatar-Russian contact situation is also desirable, in particular on the structural relationship between the code-mixing in Tatar-preferred style and the code-switching found in other “mixed” Tatar-Russian styles. A thorny obstacle to be surmounted is the difficulty of recording code-switched conversations: due to audience design and style-shifting requirements, the presence of both the fieldworker and her recording equipment prevents code-switched performance by the politically and culturally active young Tatars cited in this dissertation. Additional research on Tatar language ideologies and indexicality in linguistic performance could test the validity of the “syncretic code” hypothesis for Tatar-preferred style, which posits that the code-mixing of Russian discourse-pragmatic words found in this style may be used to construct an identity that is desirable to young urban Tatars, that of bilingual speakers who are comfortable in both the Tatar and Russian worlds but choose to live in Tatar.

Additional work is also required to map out more precisely the relationship between structural and lexical borrowing of the sort found in the contracting languages described here and language attrition. One goal of endangered language theory is to find structural innovations that are inherently symptomatic of language death, innovations that would not be found in a “healthy” language. Organizing and commenting upon minority-language discourse with a dominant-language metalinguistic code does indeed seem to be a form of language attrition, where the dominant language is making inroads in a linguistic functional domain rather than a social functional domain. This dominant-language discourse organization appears to be one sign of linguistic “tip” (Rouchdy 1989) in favor of the dominant language and another means of contraction of a minority language’s stylistic options. Myers-Scotton suggests that code switching is an “important mechanism for ‘deep borrowing’ and shift phenomena” and is “involved in language death” (1992: 33) – here, “the dying language need not have *lost* features so much as substituted those of the ‘invading’ language” (ibid. 52). The Tatar of young bilinguals, even the most ideologically devoted and competent speakers, seems to demonstrate a preference for Russian-style constructions over “characteristically Tatar” constructions, thus leading to the reduction of stylistic options. For example, if Russian *potomu chto* ‘because’ becomes the preferred means of expressing causality, or causes Tatar *chönki* ‘because’ to become the preferred construction since it is syntactically identical to the Russian, it will be at the expense of the other 27 Tatar causative constructions and aid in the loss of semantic and register distinctions. Moreover, youth Tatar in urban environments has no slang or characteristic “youth speech” innovations: Russian slang and interjections, many of which are discourse-pragmatic in function, already fill that gap.

Errington (1998: 98-100) notes that there are various approaches to the study of language contact and contact phenomena. Weinreich (1953) focussed on individual bilingualism,

in particular, on the ways that the grammar of a first language shapes the acquisition of a second language and causes “interference.” For Weinreich the locus of contact is the individual. Thomason and Kaufman (1988) moved the locus of study from the individual to groups in contact that were broadly contextualized both socially and historically. Their focus is on “borrowing,” both lexical and structural, and on the correspondence between intensity of pressure from a dominant group and amount of linguistic influence found in the minority language. Errington labels Weinreich’s approach “cognitive” and Thomason and Kaufman’s “collective.” His own fieldwork led him to focus on another locus of language contact: face-to-face interpersonal interactions that index the broader socio-cultural context. In this dissertation, I have used all three of these approaches and loci of language contact: the individual is the focus in the analysis of style-shifting, which is one form of intra-speaker variation. Style-shifting in turn is motivated and explained by audience design, a theory based on the study of interpersonal interactions. Additionally, the Tatar data that forms the basis of my analysis comes from interpersonal interactions in a variety of channels, both oral and written. These interpersonal interactions are always understood within the larger socio-cultural and political context within which they take place.

In conclusion, the examination and analysis of Tatar in post-Soviet Tatarstan demonstrates that social factors – often given only the most cursory treatment in studies of language contact – are integral to understanding the motivations for and pathways of contact-induced change. Tatar linguistic performance and the Russian-influenced structural innovations found in this performance can only be understood by analyzing all levels of language contact: intraspeaker variation and style shifting, interpersonal interactions, and higher-level social structures, ideologies, and practices, in particular, Tatar nationalism and the discourse of purity that is an integral part of this nationalism.

Bibliography

- Abraham, Werner. 1991. *Discourse particles: Descriptive and theoretical investigations on the logical, syntactic and pragmatic properties of discourse particles in German*. Amsterdam: Benjamins.
- Alimov, Gaiaz and Maksim Iusin. 1999. *Iazychniki* (Heathens). *Izvestia*. September 23.
- Anderson, Benedict R. 1987. *Imagined communities: Reflections on the origin and spread of nationalism*. London/New York: Verso.
- Anonymous. 1999. *Tatar bulu bik awir* (It's very difficult to be a Tatar). *Shähri Kazan*, August 6, p. 4.
- Anonymous. 2000. *Üzebezän tora* (It depends on us). *Mädäni Jomga*, January 28, p. 11.
- Äkhmät'yanov, Rifkat'. 1999. *Kurkinich!* (Scary!). *Mädäni Jomga*, November 26, p. 14.
- Äkhmätjanov, Mäxmüt. 1994. *Shovinistlar berläshä, ä bez haman yokläybix* (Chauvinists are gathering, and we are always sleeping). *Shähri Kazan*, September 27, p. 1.
- Äkhmätjanov, Mäxmüt. 1995. *Milli üzangni uyata* (The national consciousness awakes). *Shähri Kazan*, November 10, p. 6.
- Äkhmätjanov, Shamil. 1997. *Ike tellek, imesh* (Bilingualism, kind of). *Shähri Kazan*, March 24, p. 5.
- Äsrarov, Nail. 1999. *Telebez där'yasi saf bulsin* (May the river of our language be pure). *Shähri Kazan*, June 29, p. 3.
- Äydi, Täüfik". 1999. *Tatarcha soyläshäbezme? Döresen yazabixmi?* (Are we speaking Tatar? Are we writing correctly?). *Mädäni Jomga*, February 12, p. 14.
- Bakhtin, Mikhail. 1981. *The dialogic imagination*. Austin: University of Texas Press.
- Banguoğlu, Tahsin. 1987. *Dil bahisleri*. Istanbul: Kubbealtı Neshriyatı.
- Baran, Dominika. 2000. The role of Russian function words in urban colloquial Uzbek. *Proceedings from the Eighth Annual Symposium about Language and Society*. Austin: Texas Linguistic Forum 44(1): 18-32.
- Baranov, A. N. and D. Payar. *Neuzheli*. In K. Kiseleva and D. Payar (eds.). *Diskursivnye slova russkogo iazyka: Opyt kontekstno-semanticheskogo opisaniia*. Moskva: MetaTekst. 300-307.
- Baranov, A. N., V. A. Plungian, and E. V. Rakhilina. 1993. *Putovoditel' po diskursivnym slovam russkogo iazyka*. Moskva: "Pomovskii i partneri".
- Baugh, John. 1993. Adapting dialectology: The conduct of community language studies. In D. R. Preston (ed.), *American dialect research*. Amsterdam: John Benjamins.
- Bauman, Richard. 1975. Verbal art as performance. *American Anthropologist* 77: 290-311.
- Bauman, Richard. 1977. *Verbal art as performance*. Rowley, MA: Newbury House.
- Bell, Alan. 1984. Language style as audience design. *Language in Society* 13, 145-204.
- Benningsen, Alexandre and Chantal Quelquejay. 1961. *The evolution of the Muslim nationalities of the U.S.S.R. and their linguistic problems*. Oxford/London: Central Asian Research Centre.
- Benningsen, Alexandre and Chantal Lemercier-Quelquejay. 1981. *Les Musulmans oubliés: L'Islam en U.R.S.S. aujourd'hui*. Paris: François Maspero.
- Besnier, Nico. 1994. The truth and other irrelevant aspects of Nukulaelae gossip. *Pacific Studies* 17:3 1-39.
- Blakemore, Diane. 1987. *Semantic constraints on relevance*. Oxford: Blackwell.
- Blom, Jan-Petter and John J. Gumperz. 1972. Social meaning in linguistic structure: Code-switching in Norway. In J. J. Gumperz and D. Hymes (eds.). *Directions in sociolinguistics*. New York: Holt, Rinehard & Winston. 407-434.

- Blommaert, Jan and Jef Verschueren. 1998. The role of language in European nationalist ideologies. In B. Schieffelin, K. Woolard, and P. Kroskrity (eds.). *Language ideologies*. New York: Oxford University Press. 189-210.
- Blommaert, Jan. 1999. The debate is open. In Jan Blommaert (ed.). *Language ideological debates*. Berlin/New York: Mouton de Gruyter. 1-38.
- Bourdieu, Pierre. 1999. *Language and symbolic power*. Cambridge: Harvard University Press.
- Briggs, Charles. 1984. Learning how to ask: Native metacommunicative competence and the incompetence of fieldworkers. *Language in Society* 13, 1-28.
- Brinton, Laurel J. 1996. Pragmatic markers in English: Grammaticalization and discourse functions. Berlin/New York: Mouton de Gruyter.
- Brown, Penelope and Stephen Levinson. 1979. Social structure, groups, and interaction. In K. R. Schere and H. Giles (eds.). *Social markers in speech*. Cambridge: Cambridge University Press.
- Brown, Roger and Albert Gilman. 1960. The pronouns of power and solidarity. In T. A. Sebeok (ed.). *Style in language*. Cambridge, MA: MIT Press. 253-276.
- Budagov, Lazar. 1869-1871. *Sravnitelnyi slovar turetsko-tatarskikh narietchii, so vklucheniem upotrebitelnieishikh slov arabskikh i persidskikh i s perevodom na russkii iazykh*. Moskva: Izdatel'stvo vostochnoi literatury.
- Cameron, D, E. Frazer, P. Harvey, M.B.H. Rampton, and K. Richardson. 1992. *Researching language: Issues of power and method*. London/New York: Routledge.
- Cameron, Deborah. 1995. *Verbal Hygiene*. London, New York: Routledge.
- Campbell, Lyle and Martha C. Muntzel. 1989. The structural consequences of language death. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press. 181-196.
- Collins, James. The Ebonics controversy in context. In Jan Blommaert (ed.). *Language ideological debates*. Berlin/New York: Mouton de Gruyter. 201-234.
- Comrie, Bernard. 1981. *The Languages of the Soviet Union*. New York: Cambridge University Press.
- Connor, Walker. 1984. *The national question in Marxist-Leninist theory and strategy*. Princeton: Princeton University Press.
- Cotter, Colleen M. 1996. *Irish on the air: Media, discourse, and minority-language development*. University of California, Berkeley dissertation.
- Coulmas, Florian. 1992. *Language and economy*. New York/Oxford: Blackwell.
- Csató, Éva Ágnes. 1998. Über die finiten Verbformen des gesprochenen Nordwest-Karaimischen. In *Turkologie Heute. Tradition und Perspektive. Materialien der Dritten Deutschen Turkologenkongferenz*, Leipzig, 4.-7. Oktober 1994. Wiesbaden: Harrassowitz. 47-53.
- Csató, Éva Ágnes. 1999a. Should Karaim be 'purer' than other European languages? *Studia Turcologica Cracoviensia*, 5: 81-89.
- Csató, Éva Ágnes. 1999b. Analyzing contact-induced phenomena in Karaim. In *Proceedings of the 25th Annual Berkeley Linguistics Society meeting*. Berkeley: BLS. 54-62.
- Csató, Éva Ágnes. 2000a. Some typological features of the viewpoint aspect and tense system in spoken North-Western Karaim. In Östen Dahl (ed.). *Tense and aspect in the languages of Europe*. Berlin: Mouton de Gruyter, 671-699.
- Csató, Éva Ágnes. 2000b. Syntactic code-copying in Karaim. In Östen Dahl & Maria Koptjevskaja-Tamm (eds.). *The Circum-Baltic languages: Their typology and contacts*. Amsterdam: John Benjamins. 265-277.
- Csató, Éva Ágnes. In press. Karaim: A high-copying language. In M. C. Jones, Edith Esch & Agnes Lam (eds.). *Convergence*. New York/Berlin: Mouton de Gruyter.

- Dancygier, Barbara. 1992. Two metatextual operators: Negation and conditionality in English and Polish. In Laura Buszard-Welcher, Lionel Wee, and William Weigel (eds.). *Proceedings of the Eighteenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: BLS. 61-75.
- Dawkins, Richard M. 1916. *Modern Greek in Asia Minor; a study of the dialects of Silli, Cappadocia and Pharasa, with grammar, texts, translations and glossary*. Cambridge: University Press.
- Dimmendaal, Gerrit. 1989. On language death in eastern Africa. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press. 13-32.
- Dorian, Nancy. 1981. *Language Death*. Philadelphia: University of Pennsylvania Press.
- Dorian, Nancy (ed.). 1989. *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press.
- Dorian, Nancy. 1994. Stylistic variation in a language restricted to private-sphere use. In Douglas Biber and Edward Finegan (eds.). *Sociolinguistic perspectives on register*. New York/Oxford: Oxford University Press. 217-232.
- Dorian, Nancy. 1998. Western language ideologies and small-language prospects. In L. Grenoble and L. Whaley, (eds.). *Endangered languages*. Cambridge: Cambridge University Press. 3-21.
- Dressler, Wolfgang. 1982. Acceleration, retardation, and reversal in language decay? In Robert Cooper (ed.). *Language spread: Studies in diffusion and social change*. Bloomington: Indiana University Press.
- Duranti, Alessandro. 1997. *Linguistic anthropology*. New York: Cambridge University Press.
- Einhorn, Barbara. 1993. *Cinderella goes to market*. Verso: London.
- Eliás-Olivares, Lucía. 1976. *Ways of speaking in a Chicano community: A sociolinguistic approach*. University of Texas, Austin dissertation.
- Erman, Britt. 1987. *Pragmatic expression in English: A study of you know, you see and I mean in face-to-face conversation*. Stockholm: Almqvist and Wiksell.
- Errington, J. Joseph. 1998. *Shifting languages: Interaction and identity in Javanese Indonesia*. Cambridge/ New York: Cambridge University Press.
- Espinal, M. Theresa. 1991. The representation of disjunct constituents. *Language* 67, 726-762.
- Faller, Helen. 2000. Transforming Soviet language ideologies into post-Soviet language policies in Tatarstan. *Anthropology of East Europe Review* 18(1): 81-86.
- Faller, Helen. 2001. The fallout of Soviet nationalities policies with respect to Tatarstan. Presented at the annual meeting of the Central Eurasian Studies Society, Madison, Wisconsin, October 13.
- Faller, Helen. 2003. *Repossessing Kazan: Nation-building after socialism in Tatarstan, Russia*. University of Michigan, Ann Arbor dissertation (in progress).
- Fäsakhov, Kasıym. 1994. *Tel saflığı – millät pak'lege* (The purity of a language is the purity of a nation). *Shähri Kazan*, December 9, p. 3-4.
- Fatiykhova, Z. 1998. *Tatarcha döres yazıyk* (Let's write Tatar correctly). *Shähri Kazan*, May 16, p. 2.
- Fättakh, Nurikhan. 2000. *Tatar tele chinnan da däülät teleme?! (Is the Tatar language in reality an official language?)*. *Mädäni Jomga*, October 30, p. 11.
- Fazıljanov, Äkhät. 1997. *Don Kikbot bulmiyk inde* (Let's not be Don Quixote already). *Shähri Kazan*, April 16, p. 3.
- Ferguson, Charles. 1959. Diglossia. *Word* 15: 325-40.
- Fishman, Joshua. 1967. Bilingualism with and without diglossia; diglossia with and without bilingualism. *Journal of Social Issues* 32: 29-38.

- Fishman, Joshua. 1972. National languages and languages of wider communication in developing countries. In *Language in socio-cultural change: Essays by Joshua A. Fishman*. Stanford: Stanford University Press.
- Fishman, Joshua. 1991. Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages. Clevedon/Philadelphia/Adelaide: Multilingual Matters Ltd.
- Frank, Allen J. 1998. *Islamic historiography and 'Bulghar' identity among the Tatars and Bashkirs of Russia*. Leiden/Boston/Köln: Brill.
- Fraser, Bruce. 1988. Types of English discourse markers. *Acta Linguistica Hungarica* 38, 19-33.
- Fraser, Bruce. 1990. An approach to discourse markers. *Journal of Pragmatics* 14, 383-95.
- Fraser, Bruce. 1996. Pragmatic markers. *Pragmatics* (6)2, 167-190.
- Fraser, Bruce. 1998. Contrastive discourse markers in English. In Andreas Jucker and Yael Ziv (eds.). *Discourse markers: Descriptions and theory*. Amsterdam/Philadelphia: John Benjamins. 301-326.
- Gal, Susan. 1979. *Language shift: Social determinants of linguistic change in bilingual Austria*. New York: Academic Press.
- Gal, Susan. 1989. Language and political economy. *Annual Review of Anthropology*, 18: 345-67.
- Galiev, Gali. 1998. *Latin älifbası telne chıstartırmi?* (Will the Latin alphabet clean the language?). *Shähri Kazan*, June 17, p. 3.
- Gallämov, S. 1998. *Äy tatarning ränjetelgän mesken "h" awazı* (O the offended, poor sound "h" of Tatar). *Shähri Kazan*, February 29, p. 3.
- Ganiev, F. Ä. 2000. *Tatarcha-Ruscha uku-ukütu süzlege* (Tatar-Russian academic dictionary). Kazan: Izdatel'stvo "Rannur."
- Geertz, Clifford. 1983. *Local knowledge: Further essays in interpretive anthropology*. New York: Basic Books.
- Gil, David. 2001. Escaping eurocentrism: Fieldwork as a process of unlearning. In Paul Newman and Martha Ratliff (eds.). *Linguistic fieldwork*. Cambridge/New York: Cambridge University Press. 102-132.
- Goffman, Erving. 1963. *Behavior in public places*. New York: Free Press of Glencoe.
- Goffman, Erving. 1981. *Forms of talk*. Philadelphia: University of Pennsylvania Press.
- Grenoble, Lenore. 1998. *Deixis and information packaging in Russian discourse*. Amsterdam/Philadelphia: John Benjamins.
- Grenoble, Lenore and Lindsay Whaley (eds.). 1998. *Endangered languages*. Cambridge: Cambridge University Press.
- Grenoble, Lenore and Lindsay Whaley. 1998. Toward a typology of language endangerment. In L. Grenoble and L. Whaley (eds.). *Endangered languages*. Cambridge: Cambridge University Press. 22-54.
- Grice, H. Paul. 1975/1989. Logic and conversation. In Peter Cole and Jerry Morgan (eds.). *Syntax and semantics 3: Speech acts*. New York: Academic Press. 41-58. Reprinted in H. Paul Grice. *Studies in the way of words*. Cambridge: Harvard University Press.
- Gumperz, John. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.
- Halliday, M.A.K. and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Haarman, Harold. 1985. The impact of group bilingualism in the SU. In I. Kreindler (ed.). *Sociolinguistic perspectives on Soviet national languages: Their past, present, and future*. Berlin: Mouton de Gruyter. 313-344.

- Hamp, Eric. 1989. On signs of health and death. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press. 197-210.
- Haugen, Einar. 1989. The rise and fall of Norwegian in America. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press. 61-73.
- Heyd, Uriel. 1954. *Language reform in modern Turkey*. Jerusalem: The Israel Oriental Society.
- Hill, Jane. 1985. The grammar of consciousness and the consciousness of grammar. *American Ethnologist* 12(4): 725-738.
- Hill, Jane. 1998. "Today there is no respect": Nostalgia, "respect," and oppositional discourse in Mexicano (Nahuatl) language ideology. In B. Schieffelin, K. Woolard, and P. Kroskrity (eds.). *Language ideologies: Practice and theory*. New York: Oxford University Press. 68-86.
- Hill, Jane and Kenneth Hill. 1980a. Metaphorical switching in modern Nahuatl: Change and contradiction. In J. Kreiman and A. E. Ojeda (eds.). *Papers from the Sixteenth Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society. 121-133
- Hill, Jane and Kenneth Hill. 1980b. Mixed grammar, purist grammar, and language attitudes in modern Nahuatl. *Language in Society* 9: 321-48.
- Hill, Jane and Kenneth Hill. 1986. *Speaking Mexicano: Dynamics of syncretic language in Central Mexico*. Tucson: The University of Arizona Press.
- Hölker, Klaus. 1991. Französisch: Partikelforschung. *Lexikon der Romanistischen Linguistik*, vol V.1. Tübingen: Niemeyer, 77-88.
- Hyman, Larry. 2001. Fieldwork as a state of mind. In Paul Newman and Martha Ratliff (eds.). *Linguistic fieldwork*. Cambridge/New York: Cambridge University Press. 15-33.
- Irvine, Judith. 1989. When talk isn't cheap: Language and political economy. *American Ethnologist* 16(2): 248-67.
- Iskändärova, N. 2001. *Min Tatar tügel, akbrisi* (I am not a Tatar, it would seem). *Watanim Tatarstan*, February 6, p. 6.
- Iskhakov, D.M and R. Musina. 1991. *Sovremennye mezhnatsional'nye protsessy v Tatatskoi SSR*. Kazan: Kazanskii Nauchnyi Tsentr.
- Izvestia Press Digest. March 6, 1998. Found at www.izvestia.ru.
- Jaffe, Andrea. 1999. Locating power: Corsican translators and their critics. In Jan Blommaert (ed.). *Language ideological debates*. Berlin/New York: Mouton de Gruyter. 39-66
- Jahangari, N. 1980. A sociolinguistic study of Tehrani Persian. University of London dissertation.
- Jälälöva, Gölfäniyä. 1997. *Tatar buluim belän gorur min* (I'm proud to be a Tatar). *Shähri Kazan*, March 22, p. 2.
 - Jälälöva, Golnar. 1996. *Bezgä nindi süz kiräk* (*Terminaribiz turında uylashböyk*) (Which word do we need? (Let's think about our terminology)). *Shähri Kazan*, October 30, p. 3.
 - Jälilöva, El'mira. 2000. *Uriš telle, Tatar küngelle Oskar* (Russian-language, Tatar-souled Oskar). *Shähri Kazan*, July 7, p. 6.
 - Johanson, Lars. 1998. The history of Turkic. In Lars Johanson and Eva Csató (eds.). *The Turkic languages*. New York/London: Routledge. 81-125.
 - Jucker, Andreas and Yael Ziv. 1998. *Discourse markers: Introduction*. In Andreas Jucker and Yael Ziv (eds.). *Discourse markers: Descriptions and theory*. Amsterdam/Philadelphia: John Benjamins, 1-12.
 - Kärimi, Ä. 1999. *Ber awaz tiräsändä* (About one sound). *Mädäni Jomga*, June 9, p. 11.

- Kärimullina, Mädinä. 2001. *Telebezñeng maturlıgın saklıyk!* (Let's save the beauty of our language!). *Watanım Tatarstan*, Feb 29, p. 5.
- Karush, Sarah. 1999a. Tatars leave Russian alphabet behind. *The Moscow Times*. October 14.
- Karush, Sarah. 1999b. Tsar of Tatarstan. *The Moscow Times*. October 23.
- Kayumov, Mökhärräm. 1998. *Yazmışına yaketi birä jjat nuri*. Shähri Kazan, August 14, p. 6.
- Kefeli, Agnès. 1999. *Neylüfer-Khanym's nifty notes on grammar*. Unpublished manuscript.
- Kefeli-Clay, Agnès. 2001. *Krasben apostasy: Popular religion, education, and the contest over Tatar identity (1856-1917)*. Arizona State University dissertation.
- Khabenskaya, E. O. 2001. *Tatarskaya intelligentsiya Saratova: Osobennosti etnicheskoy samoidentifikatsii*. *Ab Imperio* 2(3): 327-354.
- Khalil, G. 2000. *My – bolgary, a ne tatory* 'We are Bulgars, and not Tatars'. Kazan.
- Khäkimov, Rafael'. 1991. *Tatarlar kemnär alar?* (Tatars – Who are they?). *Shähri Kazan*, May 8, p. 3.
- Khäyrullin, M. 1998. *Tellärne gamälgü ashiru komitetında* (On the committee for the implementation of languages). *Mädäni Jomga*, January 30, p. 11.
- Khäyrullina, Läylä. 1999. *Tatarcha uku – gorurlık ta, batırlık ta* (Studying in Tatar – it is both pride and heroism). *Mädäni Jomga*, September 24, p. 12-13.
- Khäyrullina, Läylä. 2001. *Kemnän kurkabız?* (Who are we afraid of?). *Tatar Yäshläre*. June 16, p. 5.
- Khujamät, Fäniyä. 1999. *Küngel küze kürmäsä...* (If the soul's eye doesn't see...). *Mädäni Jomga*, April 16, p. 18.
- King, Ruth. 1989. On the social meaning of linguistic variability in language death situations: Variation in Newfoundland French. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press. 139-148.
- Kirkwood, Michael (ed.). 1989. *Language planning in the Soviet Union*. London: MacMillian Press, Ltd.
- Kirkwood, Michael. 1989. Language planning: Some methodological preliminaries. In M. Kirkwood (ed.). *Language planning in the Soviet Union*. London: MacMillian Press, Ltd. 1-22.
- Kisleva, K. L. 1998. *Konechno*. In K. Kiseleva and D. Payar (eds.). *Diskursivnye slova russokogo iazyka: Opyt kontekstno-semanticheskogo opisaniia*. Moskva: MetaTekst. 345-355.
- Kondrashov, Sergei. 2000. *Nationalism and the drive for sovereignty in Tatarstan, 1988-92*. New York: St. Martin's Press.
- Kotlikäläm, Imammalik. 1997. *Bäylänchek süzlärdän arınıyk!* (Let's get rid of intrusive words!). *Shähri Kazan*, March 15, p. 3.
- Krauss, Michael. 1992. The world's languages in crisis. *Language* 68: 4-10.
- Kreindler, Isabelle. 1982. The changing status of Russian in the Soviet Union. *International Journal of the Sociology of Language*. 33: 7-39.
- Kreindler, Isabelle. 1985. *Sociolinguistic perspectives on Soviet national languages*. Berlin: Mouton de Gruyter.
- Kreindler, Isabelle. 1995. Soviet Muslims: Gains and losses as a result of Soviet language planning. In Yaacov Ro'i (ed.). *Muslim Eurasia: Conflicting legacies*. London: Frank Cass.
- Kroskrity, Paul. V. 1998. Arizona Tewa Kiva speech as a manifestation of a dominant language ideology. In B. Schieffelin, K. Woolard, and P. Kroskrity (eds.). *Language ideologies*. New York: Oxford University Press. 103-122.
- Kurbatov, Khälif. 2000. *Tatar alfavitı häm orfografıyäse* 'The Tatar alphabet and orthography'. *Miras* 2: 41-47.

- Kuter, Lois. 1989. Breton vs. French: Language and the opposition of political, economic, social, and cultural values. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press. 75-90.
- Labov, William. 1972. Some principles of linguistic methodology. *Language in Society* 1, 97-120.
- Lewis, Geoffrey. 1968. *Turkish grammar*. Oxford/New York: Oxford University Press.
- Lewis, Geoffrey. 1999. *The Turkish language reform: A catastrophic success*. Oxford: Oxford University Press.
- Lewis, Julie. 2002. *Social influences on female speakers' pitch*. University of California, Berkeley dissertation.
- Lotfullin, Marat. 2000. *Tugan teldä belem aluga shartlar tudüru kürsätkeche* (An index of the establishment of conditions for learning in the native language). *Fän Häm Tel* 4:54-55.
- Luchshikb skbol Kazani. 1996. Kazan: Khoter.
- Maksimova, N. 2000. *Tatarcha sorıybız, ruscha jawap alabız* (We ask in Tatar, and receive an answer in Russian). *Shähri Kazan*, March 16, p. 2.
- Marshall, David F. 1996. A politics of language: Language as a symbol in the dissolution of the Soviet Union and its aftermath. *International Journal of the Sociology of Language* 118: 7-42.
- Maschler, Yael. 1994. Metalanguaging in bilingual conversation. *Language in Society* 23, 325-366.
- Matsuzato, Kimitaka. From ethno-Bonapartism to centralized *caciquismo*: Characteristics and origins of the Tatarstan political regime, 1990-2000. *Journal of Communist Studies and Transition Politics*. 17(4): 43-77.
- Mazhar, Mökhämmät. 1996. "H" awazına bälakät yanıy (The letter 'h' burns and perishes). *Shähri Kazan*, June 2, p. 2.
- Mädriyeva, Monjiyä. 1998a. "Shura" kingesh digän süz ("Shura" is a word of advice). *Mädäni Jomga*, September 25, p. 13.
- Mädriyeva, Mäniyä. 1998b. *Süzlär küp kiräk, törlese kiräk* (Many and various words are necessary). *Shähri Kazan*, June 24, p. 13.
- Mäjitov, M. 1996. *Telebezne, dinebezne saklıyk!* (Let's save our language and our religion!). *Shähri Kazan*, December 7, p. 3.
- Mäkhmütov, M. I. 1993. *Tatar ädäbi telenä kergän garäp-farsä elementları*. In M. I. Mäkhmütov, K. Z. Khämzin, and G. Sh. Säyfullin (eds.). 1993. *Garäpchä-Tatarcha-Ruscha Alınnmalar Süzlege*. Kazan: "Iman" Näshriyatı, p. 795-853.
- Mäkhmütov, M. I., K. Z. Khämzin, and G. Sh. Säyfullin. 1993. *Garäpchä-Tatarcha-Ruscha Alınnmalar Süzlege*. Kazan: "Iman" Näshriyatı.
- Merkulova, M. V. 1998. *Chto li*. In K. Kiseleva and D. Payar (eds.). *Diskursivnye slova russokogo iazyka: Opyt kontekstno-semanticheskogo opisaniia*. Moskva: MetaTekst. 336-345.
- Milroy, L., W. Li and S. Moffat. 1991. Discourse patterns and fieldwork strategies in urban settings: Some methodological problems for researchers in bilingual communities. *Journal of Multilingual and Multicultural Development*. 12:4, 287-300.
- Milroy, Lesley. 1987a. *Language and social networks*, 2nd edition. Oxford/New York: B. Blackwell.
- Milroy, Lesley. 1987b. *Observing and analysing natural language: A critical account of sociolinguistic method*. Oxford/New York: Blackwell.
- Minhaj, Ämin. 1993. *Tatar telenä – tärräkıyat* (For the Tatar language – progress). *Shähri Kazan*, March 24, p. 2.
- Minhaj, Ämin. 1996. *Tugan teldä – millät yazmıshı* (In a native language – the fate of a nation). *Shähri Kazan*, November 30, p. 3.

- Minhaj, Ämin. 1997. *Sözne alıştırü –Bik jıtdı yışh* (The replacement of a word is very serious work). *Shähri Kazan*, September 3, p. 2.
- Mirsäyetov, Fäiz. 1997. *Möstäküyl дәülärkä üz kbäräfläre* (For an independent government, its own alphabet). *Shähri Kazan*, September 6, p. 3.
- Mirsäyetov, Faiz. 1999. *Tatar ilendäge kiosklarda nigä tatar gazetalari yuk?* (Why are there no Tatar newspapers in the kiosks of a Tatar country?). *Shähri Kazan*, February 10, p. 3.
- Mirzayanov, Vil. 1998. *Tatarlık belän gorurlanu gına jıtmi, anıng mänfägat'lären jing sizganıp yaklarga da kiräk* (Pride in Tatariness isn't sufficient, it's also necessary to roll up your sleeves and get to work supporting its interests). *Shähri Kazan*, May 27, p. 3.
- Mithun, Mariane. 1989. The incipient obsolescence of polysynthesis: Cayuga in Ontario and Oklahoma. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press. 243-258.
- Monguš, D. N. and S. C. Sat. 1969. *Tuvınsküi iazyk*. In N. A. Baskakov (ed.). *Tıurskie, finno-ugorskie i mongol'skie iazyki*. Moscow.
- Moro, S. 1998. *Kstati*. In K. Kiseleva and D. Payar (eds.). *Diskursivnye slova russokogo iazyka: Opyt kontekstno-semanticheskogo opisaniia*. Moskva: MetaTekst. 248-256.
- Moskovich, Wolf. 1989. Planned language change in Russian since 1917. In M. Kirkwood (ed.). *Language planning in the Soviet Union*. London: MacMillian Press, Ltd.
- Mostafin, Rafael'. 1999. *Öch mengenge yilda tatar milläte saklanırmi?* 'In the third millenium, will the Tatar nation be preserved?' *Shähri Kazan*, June 18, p. 3.
- Möhibullin, Nägäym. 1995. *Din belän tel bezne saklap kalgan* (Religion and language preserved us). *Shähri Kazan*, December 29, p. 2.
- Möstafin, R. Ä. and Ä. Kh. Khäsänov, 1995. *Mintimer Shäymiyev: Tatarstannıng berenche prezidentı*. (Mintimer Shäymiyev: First president of Tatarstan). Kazan: Tatarstan Kitap Näshriyatı.
- Mustafin, M. R. and R. G. Khyzeyev. 1994. *Vsyë o Tatarstane: Ekonomiko-geographicheskii spravochnik*. Kazan': Tatarskoe Knizhnoe Izdatel'stvo.
- Myers-Scotton, Carol. 1992. Code-switching as a mechanism of deep borrowing, language shift, and language death. In M. Brenzinger (ed.). *Language death: Factual and theoretical explorations with special reference to East Africa*. Berlin/New York: Mouton de Gruyter. 31-58.
- Myers-Scotton, Carol. 1993. *Duelling languages: Grammatical structure in code-switching*. Oxford: Clarendon Press.
- Myers-Scotton, Carol. 1998. A way to dusty death: The Matrix Language turnover hypothesis. In L. Grenoble and L. Whaley (eds.). *Endangered languages*. Cambridge: Cambridge University Press. 289-316.
- Nizamov, Ildar. 2001. *Anam tele – janım tele* (My mother tongue is the language of my soul). *Zaman*, January 6, p. 6.
- Nogmani, Rämzi. 1991. *Tatarcha belem aliyk* (Let us learn in Tatar). *Shähri Kazan*, August 10, p. 3.
- Ochs, Elinor. 1988. *Culture and language development: Language acquisition and language socialization in a Samoan village*. Cambridge: Cambridge University Press.
- Olshtain, Elite and Shoshana Blum-Kulka. 1989. Happy Hebrish: Mixing and switching in American-Israeli family interactions. In Susan Gass et al. (eds.). *Variation in Second Language Acquisition*. Clevedon/Philadelphia: Multilingual Matters Ltd. 59-83.
- Ozolins, Uldis. 1996. Language policy and political reality. *International Journal of the Sociology of Language* 118. 181-200.

- Östman, Jan-Ola. 1981. *You know: A discourse functional approach*. Amsterdam: John Benjamins.
- Pfaff, Carol. 1979. Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language* 55: 291-318.
- Poplack, Shana. 1993. Variation theory and language contact. In D. R. Preston (ed.). *American dialect research*. Amsterdam: John Benjamins. 251-286.
- Ramazanov, Sh. 1954. *Tatar tele buencha ocherklar* (Sketches of the Tatar language). Kazan': Tatknigozdat.
- Ramazanov, R. 1998. *Telebez öchen dä, määdäniyätebez öchen dä izge esb.* (Holy work on behalf of our language and our culture). *Shähri Kazan*, May 29, p. 3.
- Rampton, M.B.H. 1992. Scope for empowerment in sociolinguistics? In D. Cameron et al. (eds.). *Researching language: Issues of power and method*. London/New York: Routledge. 29-64.
- RFE/RL. 2002. The Radio Free Europe/Radio Liberty Tatar-Bashkir report. Located on the internet at: <http://reports.rferl.org>.
- Richards, Julia Becker. 1989. Mayan language planning for bilingual education in Guatemala. *International Journal of the Sociology of Language* 77. 93-115.
- Rickford, John and Faye McNair-Knox. 1994. Addressee- and topic-influenced style shift: A quantitative sociolinguistic study. In D. Biber and E. Finegan (eds.). *Sociolinguistic perspectives on register*. New York/Oxford: Oxford University Press. 235-277.
- Romaine, Suzanne. 1994. *Language in society*. Oxford: Oxford University Press.
- Romaine, Suzanne. 1995. *Bilingualism*. Oxford: Blackwell Publishers.
- Rorlich, Azade-Ayshe. 1986. *The Volga Tatars: A profile in national resilience*. Stanford: Hoover Institution Press.
- Rouchdy, Aleya. 1989. "Persistence" or "tip" in Egyptian Nubian. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press. 91-102.
- Rouchota, Villy. 1998. Procedural meaning and parenthetical discourse markers. In Andreas Jucker and Yael Ziv (eds.). *Discourse markers: Descriptions and theory*. Amsterdam/Philadelphia: John Benjamins. 97-126.
- Sadriyev, Azat. 1997. *Min nigä tatarcha ukirga buldim?* (Why did I study in Tatar?). *Shähri Kazan*, September 20, p. 2.
- Safiullin, Äkhät. 1999. *Belmim inde* (I don't know). *Watanim Tatarstan*, August 10, p. 3.
- Safiullina, Flyora and El'vira Fyodorova. 2000. *Khäzerge tatar ädäbi telendä garäp-farsä alinmalarining kire kaytu protsessi* (The process of the return of Arabic-Persian borrowings in the present-day Tatar literary language). *Fän häm Tel* No 3, 16-21.
- Sälimov, Zöhär. 2001. *Süzanna – redaktsiyäbez kunagi*. *Tatar Ile*, 15(428), p. 2.
- Schermerhorn, R.A. 1970. *Comparative ethnic relations: A framework for theory and research*. New York: Random House.
- Schiffman, Harold. 1996. *Linguistic culture and language policy*. London/New York: Routledge.
- Schiffman, Deborah. 1987. *Discourse markers*. Cambridge/New York: Cambridge University Press.
- Schilling-Estes, Natalie. 1998. Investigating 'self-conscious' speech: The performance register in Ocracoke English. *Language in Society* 27, 58-83.
- Schourup, Lawrence. 1985. *Common discourse particles in English conversation*. New York: Garland.
- Schourup, Lawrence. 1999. Discourse markers. *Lingua* 107, 227-265.

- Schmidt, Annette. 1991. Language attrition in Boumaa Fijian and Dyirbal. In Herbert Seliger and Robert Vago (eds.). *First language attrition*. Cambridge: Cambridge University Press. 113-124.
- Schur, Nathan. 1995. *The Karaite encyclopedia*. Frankfurt am Main: Peter Lang.
- Searle, John. 1969. *Speech Acts*. Cambridge: Cambridge University Press.
- Shäfi, Riza. 1992. *Il kürke – tel, tel kürke – süç* (The beauty of a country – language, the beauty of a language – the word). *Shähri Kazan*, April 22, p. 3.
- Shamsutdinova, Rasima. 2001. *Meditsinskaya terminologiya v Tatarskom yazyke* (Medical terminology in Tatar). Kazan: Regent’.
- Shäykiyev, Galim. 1993. *Isemeng kem sineng?* (Whose name is your name?). *Shähri Kazan*, October 16, p. 3.
- Shäymiyev, Mintimer. 1999. *Tatarstan Respublikasi Zakoni: Latin grafikasi nigezendä Tatar alfavitin torğışu turında* (Law of the Republic of Tatarstan: About the construction of the Tatar alphabet with a basis in Latin graphics). Law passed September 15, 1999: cited as printed in *Mädäni Jomga*, November 26, p.12.
- Shloush, Shelley. 1998. A unified account of Hebrew *bekicur* ‘in short’: Relevance theory and discourse structure considerations. In Andreas Jucker and Yael Ziv (eds.). *Discourse Markers: Descriptions and theory*. Amsterdam/Philadelphia: John Benjamins. 61-82.
- Shnirelman, Victor A. 1996. *Who Gets the Past? Competition for ancestors among non-Russian intellectuals in Russia*. Washington, D.C.: The Woodrow Wilson Center Press; Baltimore: Johns Hopkins Press.
- Silverstein, Michael. 1979. Language structure and linguistic ideology. In R. Clyne, W. Hanks, and C. Hofbauer (eds.). *The elements: A parasession on linguistic units and levels*. Chicago: Chicago Linguistics Society, 193-247.
- Sperber, Dan and Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Oxford: Blackwell.
- Spitulnik, Debra. 1998. Mediating unity and diversity: The production of language ideologies in Zambian broadcasting. In B. Schieffelin, K. Woolard, and P. Kroskrity (eds.). *Language ideologies*. New York: Oxford University Press. 163-188.
- Suny, Ronald Grigor. 1993. *The revenge of the past: nationalism, revolution, and the collapse of the Soviet Union*. Stanford: Stanford University Press.
- Sweetser, Eve. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects*. Cambridge: Cambridge University Press.
- Tankut, Hasan R. 1963. *Atatürk’ün dil çalışmaları. Atatürk ve türk dili*. Ankara: TDK. 111-36.
- Tatarsko-Russkii Slovar’*. 1996. Moskva: Izdatel’stvo Entsiklopediia.
- Taylor, Allan. 1989. Problems in obsolescence research in the Gros Ventre of Montana. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press. 167-180.
- Thomason, Sarah and Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley/Los Angeles: University of California Press.
- Traugott, Elizabeth Clauss. 1982. From propositional to textual and expressive meanings; Some semantic aspects of grammaticalization. In W. P. Lehmann and Y. Malkiel (eds.). *Perspectives on historical linguistics*. Amsterdam: Benjamins.
- Traugott, Elizabeth Clauss. 1989. On the rise of epistemic meanings in English. *Language* 63: 1, 31-55.
- Traugott, Elizabeth Clauss. 1995. The role of the development of discourse markers in a theory of grammaticalization. Paper presented at the Twelfth International Conference on Historical Linguistics, Manchester. (Revised in November 1997).

- Trudgill, Peter. 1972. Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society* 1: 179-95.
- Tsitsipis, Lukas. 1989. Skewed performance and full performance in language obsolescence: The case of an Albanian variety. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press. 117-137.
- Urla, Jacqueline. 1987. *Being Basque, speaking Basque: The politics of language and identity in the Basque country*. University of California, Berkeley dissertation.
- Usmanov, Mödärris. 1997. *Telebezgä chistalik ingsen* (May purity sink into our language). *Shähri Kazan*, October 31, p.6.
- Väliyev, Foat 1999. *Tugan telne tugaru, yaki terminologik bashbashtaklik* (The unharnessing of the native tongue, or the tyranny of terminology). *Mädäni Jomga*, June 18, p. 9.
- Väliyev, Foat. 2001. *Fänni telme, ällä uram teleme?* (Scholarly language, or street language?). *Shähri Kazan*, January 20, p. 2.
- Väliyev, Razil 1992. *Telebez khasta kbälendä* (Our language is in a sick state). *Shähri Kazan*, June 20, p. 3.
- Verschueren, Jef. 1999. *Understanding pragmatics*. London/New York: Arnold.
- Versteegh, Kees. 1997. *The Arabic language*. New York: Columbia University Press.
- Wagizov, S. G. and R. G. Wälitova. 2000. *Latin grafikası nigezendä älifba*. Kazan: Mägarif Näshriyatı.
- Walker, Edward W. 1996. The dog that didn't bark: Tatarstan and asymmetrical federalism in Russia. *The Harriman Review* (9)4.
- Watson, Seosamh. 1989. Scottish and Irish Gaelic: The giant's bed-fellows. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. Cambridge: Cambridge University Press. 41-59.
- Weinreich, Uriel. 1953. *Languages in contact*. New York: Columbia University Press.
- Wolfson, Nessa. 1976. Speech events and natural speech: Some implications for sociolinguistic methodology. *Language in Society* 5(2), 189-210.
- Woolard, Kathryn. 1989. *Double talk: bilingualism and the politics of ethnicity in Catalonia*. Stanford: Stanford University Press.
- Woolard, Kathryn and Bambi Schieffelin. 1994. Language ideology. *Annual Review of Anthropology*, 23: 55-82.
- Yurchak, Alexei. 2000. Privatize your name: Symbolic work in a post-Soviet linguistic market. *Journal of Sociolinguistics* 4 (3): 406-434.
- Zajaczkowski, Ananiasz. 1961. *Karaims in Poland: History, language, folklore, science*. Warsaw: Panstwowe Wydawnitwo Naukowe.
- Zakiev, Mirfatikh Zakiyevich (ed.). 1995. *Tatarskaia Grammatika: Sintaksis (Tom III)*. Kazan: Akademiya Nauk Tatarstana.
- Zakiev, Mirfatikh Zakiyevich (ed.). 1997. *Tatarskaia Grammatika: Morfologiya (Tom II)*. Kazan: Akademiya Nauk Tatarstana.
- Zakiev, Mirfatikh Zakiyevich. 2000. *Tatar-Törki Etnogenezı* (Tatar-Turkic Ethnogenesis). Kazan.
- Zinnätullin, Gabbaz. 2000. *Minem uris kızı bulasım kilä!* (I want to be a Russian girl!). *Watanım Tatarstan*, March 28. p. 3.
- Zolotov, Andrei (Jr.) 1999. Keeping Islam in line. *The Moscow Times*. May 16.

Appendix A

Transliteration conventions and grammatical abbreviations

Transliteration Conventions

The following transliteration conventions are used in this dissertation:

Tatar (The same transliteration system is used to represent Karaim and Turkish)

Transliteration	Cyrillic	Phoneme	Transliteration	Cyrillic	Phoneme
a	а	/a/	m	м	/m/
ä	ә	/æ/	n	н	/n/
b	б	/b/	ng	ң	/ŋ/
ch	ч	/tʃ/	o	о	/o/
d	д	/d/	ö	ө	/ø/
e	е	/e/	p	п	/p/
f	ф	/f/	r	р	/r/
g	г	/g/	s	с	/s/
g''	гъ	/ɣ/	sh	ш	/ʃ/
h	һ	/h/	t	т	/t/
i	и	/i/	u	у	/u/
ï	ы	/i/	ü	ү	/y/
j	ж	/dʒ/	v	в	/v/
k	к	/k/	w	в	/w/
k''	къ	/q/	y	й	/j/
kh	х	/x/	z	з	/z/
l	л	/l/			

Russian

Transliteration	Cyrillic	Phoneme	Transliteration	Cyrillic	Phoneme
a	а	/a/	n	н	/n/
b	б	/b/	o	о	/o/
ch	ч	/tʃ/	p	п	/p/
d	д	/d/	r	р	/r/
e	е	/e/	s	с	/s/
f	ф	/f/	sh	ш	/ʃ/
g	г	/g/	shch	щ	/ʃ/
h	һ	/h/	t	т	/t/
i	и	/i/	ts	ц	/ts/
ï	й	/j/	u	у	/u/
k	к	/k/	v	в	/v/
kh	х	/x/	y	ы	/i/
l	л	/l/	z	з	/z/
m	м	/m/			

Grammatical Abbreviations

1PP	first person plural	IMP	imperative
1PS	first person singular	INDEF.PART	indefinite particle
2PP	second person plural	INF	infinitive
2PS	second person singular	INST	instrumental
3PP	third person plural	INT.PART	interrogative particle
3PS	third person singular	LOC	locative
?PT	interrogative particle	NEG	negative
1 st GER	first gerund in Tatar	ORD	ordinal
2 nd GER	second gerund in Tatar	PART	particle
A.NONPST	Karaim non-past verb in –A	PAST	past tense
ABL	ablative	PAST.DEF	definite past tense
ACC	accusative	PAST.PART	past participle
ADV	adverb	PERF.PAST	past tense, perfective aspect
AGT	agentive	PL	plural
AOR	aorist	POL	politeness suffix
CAUS	causative	POSS	possessive suffix
COMP	comparative	PRES	present tense
COND	conditional	R.NONPAST	Karaim non-past verb in –r
CONTPST	continuous past tense	REL.PRO	relative pronoun
DAT	dative	RFL	reflexive
DIST	distributive	SIMPPST	simple past tense
EMPH.PART	emphatic particle	SUBJ	subjunctive
ENC	enclitic	SUPER	superlative
FUT	future	V.NOUN	verbal noun
FUT.DEF	definite future tense		
GEN	genitive case		
GER	gerund		

Appendix B

Russian discourse-pragmatic words code-mixed in Tatar

<i>Code-mixed Russian word</i>	<i>Tatar equivalent</i>	<i>Morphosyntactic realization</i>
<i>Discourse structure marking – subordinating</i>		
chto ‘that’	<ol style="list-style-type: none"> 1. Verbal noun in –u, possessed and declined like a noun. 2. gan + ni with subject in absolute case. 3. gan + possessive endings + accusative. 4. asi in accusative. 5. achak, possessed and in accusative. 6. achak in accusative with an absolute subject. 7. dip to demarcate the end of the subordinate clause. 8. ki to demarcate the beginning of the subordinate clause. 9. subordinate clause with verb in passive and demonstrative pronoun in absolute case. 10. subordinate clause with demonstrative pronoun in locative case. 11. demonstrative pronoun in accusative 	<p>Utterances with <i>chto</i> have a significantly different morphosyntactic realization than those using any of the 11 options to the left. A Tatar sentence using <i>chto</i> will have a subordinate clause headed by <i>chto</i> following the main clause, both with finite verbs. Constructions 1-6 use non-finite verbs in the subordinate clause, which is embedded immediately preceding the main verb. Constructions 7-8 use finite verbs in both clauses, and have the subordinate clause embedded immediately preceding the main verb. Constructions 9-11 use finite verbs in both clauses, and the subordinate clause, while preceding the main clause, is not embedded.</p>
chtob(y) ‘in order that, that’	<ul style="list-style-type: none"> -u öchen -rga dip -uına -sın 	<p>Completely different clausal structure and construction. When <i>chtoby</i> heads a subordinate clause, the subordinate clause follows the main clause and contains a finite verb. The Tatar constructions all have the subordinate clause embedded and preceding the main verb, which is sentence-final, and the verb of the subordinate clause is not finite (it is either a verbal noun in –u, an infinitive, a possessed verbal noun in the dative case, or there is no verb in the subordinate clause and the main clause verb is in the optative form).</p>

<i>Code-mixed Russian word</i>	<i>Tatar equivalent</i>	<i>Morphosyntactic realization</i>
potomu chto 'because'	<p>Synthetic:</p> <ol style="list-style-type: none"> 1. Y <i>bulganlıktan</i>, X. 2. Y-<i>u arkasında</i>, X. 3. Y-<i>u nätiyäsendä</i>, X. 4. Y-<i>ganlıktan</i>, X. 5. Y-<i>ganga</i>, X. 6. Y-<i>ganga kürä</i>, X. 7. Y-<i>gan öchen</i> X. 8. Y-<i>gan säbäple</i>, X. 9. Y-<i>u säbäple</i>, X. 10. Y-<i>magach</i> X-<i>ma</i>. <p>Analytic:</p> <ol style="list-style-type: none"> 1. X, <i>chönki</i> Y. 2. Y, <i>shunga kürä</i> X. 3. Y, <i>shunga</i> X. 4. Y, <i>shuning öchen</i> X. 5. Y, <i>shuning nätiyäsendä</i> X. 6. Y, <i>shul säbäple</i>, X. 7. Y, <i>shul säbäpten</i>, X. 8. Y, <i>shunlıktan</i> X. 9. Y, <i>shuning arkasında</i> X. 10. X:Y, Y:X. (written only.) 11. ...<i>menä närsägä</i> X: Y. 12. ... <i>menä nigä</i> X: Y. 13. ...<i>menä närsädän</i> X: Y. 14. ...<i>menä nidän</i> X: Y. 15. ...<i>menä ni öchen</i> X: Y. 16. ...<i>menä ni säbäptän</i> X:Y. 17. ...<i>menä närsägä kürä</i>X:Y. 18. ...<i>menä närsä arkasında</i> X: Y. 19. ...<i>menä närsä nätiyäsendä</i> X:Y. 	When used in place of <i>chönki</i> , there is no difference in the realization of the sentence. However, a sentence with <i>potomu chto</i> is significantly different from the 27 other possible constructions, particularly the ten synthetic constructions, none of which use finite verbs.
tak chto 'so'	shulay bulgach	none
<i>Discourse structure marking – coordinative</i>		
i 'and'	häm belän da...da	Russian <i>i</i> and Tatar <i>häm</i> are used identically. Tatar <i>belän</i> is more like 'with' than 'and,' but syntactically identical. The enclitic <i>da</i> comes after both nouns or verbs being coordinated.
tozhe 'also'	da shulay uk tagın	When used in place of the enclitic <i>da</i> , a slight difference in word order: <i>tozhe</i> is used only once, <i>da</i> comes after both nouns or verbs being coordinated.

<i>Code-mixed Russian word</i>	<i>Tatar equivalent</i>	<i>Morphosyntactic realization</i>
<i>Discourse structure marking – contrastive</i>		
ili ‘or’	yä yäisä yäki ällä... yä bulmasa	same
no ‘but, however’	ämma läkin	same
a to ‘else’	yuksa, yugiysä	same
zato ‘but on the other hand’	ämma, shulay da	same
<i>Metacommentary and deixis</i>		
to est’ ‘i.e.’	yag”ni döresläp äytkändä	same
koroche ‘in brief’	kiskasi	same
kstati ‘by the way’	ber ungayadan süz ungaenda	The Tatar equivalents are phrases, but otherwise no effect on overall clausal structure of the sentence
vot ‘there’	menä	same
nu ‘so’	none	n/a
naprimer ‘for example’	mäsälän	same
prichem ‘moreover’	östäwenä (tagin)	same
tem bolee ‘especially as’	bigräk tä	same
konechno	älbättä	same
chto li ‘or something,’	mï	Placement of the Tatar question particle can vary, although it is usually phrase-final: the Russian particle is always sentence-final.
neuzheli ‘really?’	(ällä) chïnlap ta ... mï	Replacement of complex morphology with simple: however, the <i>mï</i> question particle is sometimes retained.
i vsyo ‘and that’s it’	häm bette	same
izvini(te) ‘pardon, sorry’	gafu it, gafu itegez	same
vot, tak ‘so, there’	none	n/a
slushai(te) ‘listen’	tïngla äle kara äle	same
davai(te) ‘Let’s go! Come on!’	äydä! äydegez!	NB: Russian ‘ayda!’ which is used for physical movement only, is the borrowed version of this Tatar verb.
<i>Interactional performatives</i>		
privyet ‘hi’	sälam	n/a

<i>Code-mixed Russian word</i>	<i>Tatar equivalent</i>	<i>Morphosyntactic realization</i>
poka 'bye'	sau bul, sau buligiz, khush	n/a
pozhaluista 'please'	zinhar zinhar öchen -chi on imperative verbs	same
izvini(te) 'pardon, sorry'	gafu it, gafu itegez	same
<i>Evaluatives</i>		
kak 'how'	nichek	same
<i>Positive evaluation</i>		
chudo 'cool' (lit. 'miracle')	no equivalent	n/a
molodyets 'well done'	äferin	n/a
prikol' 'how amusing'	no equivalent	n/a
tochno 'precisely, exactly'	tögäl tögäl itep döres itep näk" chinnan da dörestän dä	n/a
voobshche klassno 'totally cool'	no equivalent	n/a
<i>Negative Evaluation</i>		
uzhas 'how awful'	kotochkich	n/a
zhalko 'it's a shame'	kizganich	n/a
<i>Hedges</i>		
chut' 'almost'	2 nd gerund + <i>yazırğa</i> (finite)	The Russian <i>chut'</i> precedes the finite verb, usually in the negative past tense. (In Russian, <i>chut'</i> in this usage requires a negative verb). The Tatar equivalent uses the 2 nd gerund form of the verb of action with the auxiliary verb <i>yazırğa</i> in finite form giving the semantics of 'almost.'
khot' 'at least, only'	ichmasam hich bulmasa	same
ni v koem sluchae 'by no means'	hichkaychan berkaychan da tügel	same
okolo 'around'	yanında tıräsändä	The preposition <i>okolo</i> precedes the nominal complement as opposed to the postpositions found to the left.

<i>Code-mixed Russian word</i>	<i>Tatar equivalent</i>	<i>Morphosyntactic realization</i>
pochti 'almost'	diärlek	Slight difference in word order: <i>pochti</i> precedes the nominal complement, <i>diärlek</i> follows it.
prosto 'simply, just, merely'	tik	same
sluchaino 'by chance'	kötmägändä iskärmägändä iylamaganda	Same overall syntax, but the Tatar equivalents are morphologically complex resembling secondary post-positions, and not adverbs.
<i>Epistemics</i>		
mozhet 'might'	mömkin	same
naverno 'probably, most likely'	bälki mögaen ikhtimal shayat'	same
<i>Evaluations of time and degree</i>		
uzhe 'already'	inde	same
skoro 'soon'	tiz	same
srazu 'immediately'	shul uk wakıtta khäzer ük shunda uk berdän	The first three Tatar options are complex constructions using the emphatic particle <i>uk</i> (essentially equivalent to Russian <i>уже</i>).
srochno 'quickly, soon, urgently'	ashigich	none
dazhe 'even'	khätta	The Tatar and Russian particles are periodically used to double-mark the word in question in the construction <i>dazhe...khätta</i> .
eshche 'more, still'	tagın da	same
slishkom 'too, overly'	bigräk chiktän tish chamasız ifrat	same
da 'yes'	äye	same
absoliutno 'absolutely'	no equivalent: <i>absoliut</i> (adj.) is a standard borrowing in Tatar	same
inogda 'sometimes'	kaywakit kaychakta kayber chakta	same
kak vseгда 'as always'	gadättä gechä	same

<i>Code-mixed Russian word</i>	<i>Tatar equivalent</i>	<i>Morphosyntactic realization</i>
obyzatel'no 'without fail'	hichiksez	same
obychno 'usually'	gadättä	same
tochno 'precisely, exactly'	tögäl tögäl itep döres itep näk" chinnan da dörestän dä	Can be slight difference in word order, with the Russian word preceding its head.
poluchaetsia '(it) turns out'	chıgu kilep chıgu	same
voobshche 'in general'	gomumän bötenläy härwakit	same

Appendix C

Arabic discourse-pragmatic words borrowed into Tatar

Words used in present-day Tatar

<i>Arabic borrowing</i>	<i>Gloss</i>	<i>Code-mixed Russian equiv.</i>
<i>Discourse structure marking</i>		
<i>Discourse structure marking – coordinative</i>		
wä	‘and’	<i>i</i>
<i>Discourse structure marking – contrastive</i>		
ämma	‘but, however’	<i>no</i>
läkin	‘but, however’	<i>no</i>
khayir	‘however, although’	<i>no</i>
khal’buki	‘whereas’	
<i>Metacommentary and deixis</i>		
älbättä	‘of course’	<i>konechno</i>
fölan, fälän	‘such and such’	
makhsus	‘especially, particularly’	
mäglüm (ki)	‘it is well known (that)’	
mäsälän	‘for example’	<i>naprimer</i>
nihayat’	finally	
wässäläm	‘and that’s it’	<i>i vse</i>
yag”ni	‘that is, i.e.’	<i>to est’</i>
<i>Interactional performatives</i>		
säläm	‘hello’	<i>privyet</i>
ässälame galäykem	‘hello’	<i>privyet</i>
märkhaba	‘hello’	<i>privyet</i>
khush	‘hello,’ ‘goodbye’	<i>privyet, poka</i>
räkhmät	‘thanks’	<i>spasibo</i>
gafu it(egez)	‘excuse me’	<i>izvini(te)</i>
rakhim it(egez)	‘welcome’	
jämägat’	‘people, folks’ (address term)	
mökhtäräm	‘respected’ (address term)	
<i>Evaluatives</i>		
<i>Positive evaluation</i>		
afärin	‘bravo! well done!’	<i>molodyets!</i>
amin!	‘amen’	

<i>Arabic borrowing</i>	<i>Gloss</i>	<i>Code-mixed Russian equiv.</i>
insha alla, inshalla	'God willing'	
<i>Epistemics</i>		
bälki	'perhaps, maybe'	<i>naverno</i>
ikhtimal	'probably'	<i>naverno</i>
mögayin	'probably'	<i>naverno</i>
mömkin	'possible, probable, likely'	<i>mozhet</i>
<i>Evaluations of time and degree</i>		
daimi	'constantly, continually'	
gomumän	'usually, in general'	<i>voobsbche</i>
fakat'	'only, just'	<i>tol'ko</i>
gajäp	'surprisingly'	
gayat'	'very'	
ifrat	'overly'	<i>slishkom</i>
kadär	'up to, until, as much as'	
khätta	'even'	<i>dazhe</i>
khususän	'especially, particularly'	
motlakan	'absolutely'	<i>absolutno</i>
tämamän	'completely'	<i>sovsem</i>

Archaic Arabic loanwords

<i>Arabic borrowing</i>	<i>Gloss</i>	<i>Code-mixed Russian equiv.</i>
<i>Discourse structure marking</i>		
<i>Discourse structure marking – subordinating</i>		
binabärin	'therefore, in view of this'	
falihaza	'and therefore'	
hakäza	'thus, in that way'	
käza	'thus, in that way'	
käzalik	'thus, in that way'	
lihazä	'therefore, on this basis'	
lizalik	'therefore, for this'	
räsmä	'thus, in that way'	
täki	'in order to'	<i>chtoby</i>
zirä	'because, since'	<i>potomu chto, tak chto</i>
<i>Discourse structure marking – coordinative</i>		
äyzan	'also, and'	<i>tozhe</i>
niz	'also, and'	<i>tozhe</i>

<i>Arabic borrowing</i>	<i>Gloss</i>	<i>Code-mixed Russian equiv.</i>
<i>Discourse structure marking – contrastive</i>		
äm	‘or’	<i>ili</i>
illyä	‘but, however’	<i>no</i>
lya äkalle	‘although’	
räg”män	‘despite’	
wäläkin	‘but, however’	<i>no</i>
wäläü	‘although, and though’	
wäli	‘but, however’	<i>no</i>
<i>Metacommentary and deixis</i>		
kyäshki	‘if only, would that’	
wä gaire zalik	‘and so on, etcetera’	
ila akhirihi	‘and so on, etcetera’	
ilkh	‘and so on, etcetera’	
wä hakäza	‘and so on, etcetera’	
hälömmä järä	‘and so on, etcetera’	
wäjhän	‘on one hand’	
gayanän	‘clear, clearer than clear’	
määlän	‘not literally’	
mäga ma fih	‘together with this’	
mäga haza	‘together with that’	
mägan	‘together with...’	
magada	‘except’	
ma siva	‘except’	
finäfsihi	‘basically’	
khölasan kälyäm	‘in brief’	<i>koroche</i>
älkhasiyl’	‘in brief’	<i>koroche</i>
ätäm	‘more precisely’	
khalänki	‘the fact of the matter is...’	
biäyikhalän	‘in any case, at any rate’	
galä köllettäk”dir	‘in any case, at any rate’	
ikhtiyatan	‘in any case’	
<i>Deixis</i>		
hazä	‘this’	
äl’an	‘now’	
älkhal’	‘now’	
khaliya	‘now’	
äl’yaum	‘today, now’	
lyakhikan	‘now’	
ämma bäg”dä	‘later, afterwards’	
bäg”dä	‘after’	
möäkhkhärän	‘later, afterwards’	
wä bägidä	‘after that, later’	
kablä	‘earlier’	
mökaddämä	‘before, formerly’	

<i>Arabic borrowing</i>	<i>Gloss</i>	<i>Code-mixed Russian equiv.</i>
därkhal'	'right away, immediately'	
äüwälän	'first of all, in the first place'	
bilakhir	'finally, in the end'	
bilakhirä	'finally, in the end'	
ma	'then, when'	
<i>Interactional performatives</i>		
täxiyä	'hi'	<i>privyet</i>
ählän wä märkhaba	'hello'	
<i>Evaluatives</i>		
nitän	'how? in what way?'	<i>kak</i>
nitäkin	'how? in what way?'	<i>kak</i>
nitäkkäm	'how? in what way?'	<i>kak</i>
<i>Positive evaluation</i>		
fäbiha wä nigamä	'wonderful, great'	
mägal'urra wä bilbanzay	'With a cry of ura! And with a cry of banzai!'	
<i>Negative Evaluation</i>		
äsäfa	'How awful!'	<i>uzhas</i>
älgayaze billa	'God forbid!'	
dad bidad	'Alas! How awful!'	<i>uzhas</i>
däriga	'O grief! O pity!'	
häyhat	'Alas! How awful!'	
khäef	'How awful!'	<i>uzhas</i>
khäifa ki	'How awful!'	<i>uzhas</i>
näguze billahi	'God forbid!'	
wa däriga!	'Oh, how awful!'	
wa wäilä!	'How awful!'	<i>uzhas</i>
wa!	'Ach'	
yahu	'Oh God!'	
yaläytä	'Oh, if only!'	
mä attäässef	'unfortunately'	
täässefän	'unfortunately'	
<i>Evaluations that are unclear without context</i>		
äkhökme lillyä	'everything is in the hands of God'	
gajäba	'Surprising! How strange!'	
lya kälyäm	'there aren't words'	
rabbi	'My God!'	
<i>Epistemics</i>		

<i>Arabic borrowing</i>	<i>Gloss</i>	<i>Code-mixed Russian equiv.</i>
lägallä	'maybe, probably'	
namömkin	'impossible, improbable'	
<i>Evaluations of time and degree</i>		
däfgatän	'immediately'	<i>srazu</i>
fäürän	'immediately'	<i>srazu</i>
filkhal'	'immediately'	<i>srazu</i>
gajälätän	'quickly, in a hurry'	<i>srochno</i>
galälgajlä	'quickly, soon'	<i>skoro</i>
karibän	'soon'	<i>skoro</i>
mötäwaziyan	'simultaneously'	
<i>Evaluations of degree</i>		
bäli	'yes, true'	<i>da</i>
bilgakes	'on the contrary'	
khasha vä khasha	'no, no; never'	<i>nyet</i>
läysä	negative particle	<i>nyet</i>
binnisbë	'comparatively'	
käännähü	'as if, as though'	
käma	'like, exactly like'	
kīyasän	'comparatively'	
nisbätän	'relatively'	
äjmagiyn	'all'	
bäg"zi	'none of..., none'	
chänd	'several'	
jämig"	'all'	
<i>Evaluations of frequency</i>		
bilkhak	'in reality'	
därkhäkiykat'	'in reality'	
sadiykana	'truly'	
äbädän	'forever'	
äbädiyän	'forever'	
daimän	'always, constantly'	
häraynnä	'always, constantly'	
härwakit	'always'	
jawidanä	'forever'	
taäbäd	'forever'	
äksäri	'in the majority of cases'	
äksäriyä	'in the majority of cases'	
biljömlä	'in general'	<i>voobshche</i>
gadätän	'usually'	<i>obychno</i>
galälgadät	'as usual'	<i>kak vseгда</i>
gorfän	'usually'	<i>obychno</i>
jabäja	'here, there, and everywhere'	
täk"lidän	'as usual'	<i>kak vseгда</i>
mähzhan	'exceptionally, only'	

<i>Arabic borrowing</i>	<i>Gloss</i>	<i>Code-mixed Russian equiv.</i>
bäg”zän	‘sometimes’	<i>inogda</i>
häray	‘each time’	
märrä	‘one time, once’	
mirar	‘repeatedly, more than once’	
mirarän	‘several times’	
bishäk	‘without a doubt’	
galälitlak	‘absolutely, generally’	<i>absolutno, voobshche</i>
lya mäkhälä	‘without a doubt’	
lya shäk	‘without a doubt’	
möttäkiyn	‘without a doubt’	
bittämam	‘completely’	<i>sovsem</i>
bitämamiha	‘completely’	<i>sovsem</i>
bikhakküy	‘completely’	<i>sovsem</i>
kyamilän	‘completely’	<i>sovsem</i>
torra	‘completely’	<i>sovsem</i>
bänagyah	‘unexpectedly’	
galälgaflä	‘unexpectedly’	
min khäyse lya yakhtäsib	‘unexpectedly’	

Appendix D

Persian discourse-pragmatic words borrowed into Tatar

Words used in present-day Tatar

<i>Persian borrowing</i>	<i>Gloss</i>	<i>Code-mixed Russian equiv.</i>
<i>Discourse structure marking – subordinating</i>		
ägär	‘if’	
chönki	‘because’	<i>potomu chto</i>
ki	subordinate clause indicator, can often be glossed in English as ‘that’	<i>chto</i>
<i>Discourse structure marking – coordinative</i>		
häm	‘and’	<i>i</i>
<i>Discourse structure marking – contrastive</i>		
gärchä	‘although’	
mägär	‘however, only’	<i>no</i>
ya	‘or’	<i>ili</i>
yakhut	‘or’	<i>ili</i>
yaki	‘or’	<i>ili</i>
<i>Evaluations of degree</i>		
hämmä	‘all, each’	
här	‘each, every’	
hich	‘nothing, none’	

Archaic Persian loanwords

<i>Persian borrowing</i>	<i>Gloss</i>	<i>Code-mixed Russian equiv.</i>
<i>Discourse structure marking – subordinating</i>		
öz baskeh	‘because’	<i>potomu chto</i>
ancheh	‘that which, that’	<i>chto</i>
chun, cho	‘because, after’	<i>potomu chto</i>
ta	‘so’	<i>tak chto</i>
taakeh	‘so, in order to’	<i>chtoby, tak chto</i>
zan, azan	‘since, from that’	
<i>Discourse structure marking – metacommentary and deixis</i>		
in, inha	‘this, these’	

<i>Persian borrowing</i>	<i>Gloss</i>	<i>Code-mixed Russian equiv.</i>
bädär	'there'	<i>vot</i>
<i>Positive evaluation</i>		
aya	'wonderful! really?'	
shabash	'bravo!'	<i>molodets</i> (loosely equiv.)
zehl	'bravo! great!'	<i>molodets</i> (loosely equiv.)
<i>Epistemics</i>		
bayäd	'maybe'	<i>mozhet</i>
<i>Evaluations of time and degree</i>		
ari	'yes, like that'	<i>da, tak</i>
bäli	'yes, like that'	<i>da, tak</i>
häman	'very quickly'	<i>srazu, srochno</i>
baaz	'again'	
hämishä	'constantly, continuously'	
häray	'each time'	
hägiz	'each time'	
tänha	'one, only'	
chänd	'several'	
asa	'like, similar'	
basaan	'like, as if'	
hänüz	'still, also, exactly'	<i>eshe</i>
bas	'sufficiently'	
bäsa, bäsi	'much, enough'	
bisiyar	'much, very'	

Appendix E

Turkish discourse-pragmatic words borrowed into Asia Minor Greek

<i>Turkish word or phrase</i>	<i>Gloss</i>	<i>Notes</i>
<i>Discourse structure marking – subordinating</i>		
chünki	‘because’	< Persian, also found in Tatar
eyer	‘if’	< Persian, also found in Tatar
ichin	‘for, on account of, because’	
ki	‘that’	< Persian, also found in Tatar
mademki	‘because, as long as’	
<i>Discourse structure marking – coordinative</i>		
hem	‘and, also’	< Persian, also found in Tatar
da	‘and’ (enclitic)	
demek	‘so, thus’	
<i>Discourse structure marking – contrastive</i>		
meyer	‘but’	< Persian, also found in Tatar
amma	‘but’	< Arabic, also found in Tatar
lakin	‘but’	< Arabic, also found in Tatar
yahut	‘or’	< Persian, also found in Tatar
yoksa	‘or’	
<i>Metacommentary and deixis</i>		
mi	interrogative particle	
o	‘that’	
bu	‘this’	
sonra	‘afterwards’	
sonradan	‘afterwards’	
bir vakit(ta)	‘once (upon a time)’	<i>vakit</i> is from Arabic
mahsus	‘expressly, in particular’	< Arabic, also found in Tatar
filân	‘and so on, such and such’	< Arabic, also found in Tatar
keshke	‘if only’	< Arabic, also found in Tatar (now archaic)
hayde	‘come, c’mon’	
bakalim	‘let’s see’	
<i>Interactional performatives</i>		
selam	‘hello’	< Arabic, also found in Tatar
<i>Positive evaluation</i>		
aferin	‘bravo! well done!’	< Arabic, also found in Tatar

<i>Turkish word or phrase</i>	<i>Gloss</i>	<i>Notes</i>
<i>Epistemics and deontics</i>		
belki	'perhaps, probably'	< Arabic, also found in Tatar
mutlaka	'without a doubt'	< Arabic, also found in Tatar (now nearly archaic)
<i>Evaluations of time and degree</i>		
birden bire	'suddenly, at once'	
yeniden	'anew, afresh'	
pek	'very'	
tamamen	'completely'	< Arabic, also found in Tatar
chep	intensifier	
daha	'still, already'	
beraber	'equal to, uniform with'	
dek	'as far as, until'	
kadar	'as many, as much as'	< Arabic, also found in Tatar
o kadar	'so, as much as that'	kadar is < Arabic
eyi	'yes, very well'	
hayır	'no'	< Arabic, also found in Tatar
hiç	'nothing, none'	< Persian, also found in Tatar
her	'every'	< Persian, also found in Tatar
tıbki	'quite like, similar to'	< Arabic

Appendix F

A crosslinguistic comparison of discourse-pragmatic borrowings and code-mixed words

Russian/Polish in Karaim	Arabic/Persian in Tatar ¹	Turkish in AMGreek	Russian in Tatar
<i>Subordinating conjunctions</i>			
ki 'that' (< Persian)	ki 'that'	ki 'that' ²	chto 'that'
–	–	–	chtoby 'in order that, that'
–	–	–	tak chto 'so'
–	ägär 'if'	eyer 'if'	–
–	chönki 'because'	chünki 'because'	potomu chto 'because'
–	–	ichin 'because, on account of'	–
–	–	mademki 'because, as long as'	–
373	<i>Coordinating conjunctions</i>		
–	wä 'and'	–	–
i 'and'	häm 'and'	hem 'and'	i 'and'
–	–	da 'and, also'	tozhe 'and, also'
–	–	demek 'so, thus'	–
<i>Contrastive conjunctions</i>			
a 'and/but'	–	–	a 'and'
–	–	–	a to 'else'
–	–	–	zato 'but on the other hand'
–	ämma 'but, however'	amma 'but, however'	no 'but, however'

¹ Not including now-archaic borrowings

² When the words in the Arabic/Persian in Tatar column and the Turkish in Asia Minor Greek column are the same, it means that the word was borrowed from Arabic or Persian into Turkish, and then borrowed once more from Turkish into Asia Minor Greek.

Russian/Polish in Karaim	Arabic/Persian in Tatar	Turkish in AMGreek	Russian in Tatar
-	läkin 'but, however'	lakin 'but, however'	
-	mägär 'but, however'	meyer 'but, however'	
-	khal'buki 'whereas'	-	-
-	khayir 'although, however'	(hayir means 'no')	-
-	gärcha 'although'	-	-
-	ya 'or'	-	ili 'or'
-	yahut 'or'	yahut 'or'	
-	yaki 'or'	-	
-	-	yoksa 'or'	
<i>Metacommentary and deixis</i>			
to 'then'	-	-	-
-	-	sonra 'afterwards'	-
-	-	sonradan 'afterwards'	-
-	-	bir vakit(ta) 'once (upon a time)'	-
nu 'well, so'	-	-	nu 'well, so'
nu to 'well then'	-	-	-
vot 'there'	-	-	vot 'there'
-	-	-	slushai(te) 'listen!'
-	-	hayde 'come, c'mon'	davai(te) 'come, c'mon, let's go!'
-	-	bakalim 'let's see'	-
-	jämogat' 'people, folks!'	-	-
-	älbättä 'of course'	elbette 'of course'	konechno 'of course'
-	mäglüm (ki) 'it is well known'	-	-
-	yag"ni 'that is, i.e.'	-	to est' 'that is, i.e.'
-	mäsälän 'for example'	-	naprimer 'for example'

Russian/Polish in Karaim	Arabic/Persian in Tatar	Turkish in AMGreek	Russian in Tatar
—	nihayat ‘finally’	—	—
—	—	—	koroche ‘in brief’
—	—	—	kstati ‘by the way’
—	wässälam ‘and that’s it’	—	i vse ‘and that’s it’
—	—	—	vot, tak ‘so there’
—	kyäshki (now archaic)	keshke ‘if only’	—
—	fälan ‘such and such’	filan ‘such and such’	—
—	makhsus ‘especially’	mahsus ‘especially’	—
—	—	mi [interrogative particle]	chto-li [interrogative particle]
—	—	—	neuzheli ‘really?’
—	—	o ‘that’	—
—	—	bu ‘this’	—
<i>Interactional performatives</i>			
—	sälam ‘hi, hello’	selam ‘hello’	privyet ‘hi, hello’
—	märkhaba ‘hello’	—	—
—	ässälame galäykem ‘hello’	—	—
—	rakhim it(egez) ‘welcome’	—	—
—	gafu it(egez) ‘pardon, excuse me’	izvini(te) ‘pardon, excuse me’	—
—	khush ‘bye’	—	poka ‘bye’
—	—	—	pozhaluista ‘please’
—	räkhmat ‘thank you’	—	spasibo ‘thank you’
—	mökhtäram ‘respected’	—	—

Russian/Polish in Karaim	Arabic/Persian in Tatar	Turkish in AMGreek	Russian in Tatar
<i>General Evaluatives</i>			
–	–	–	kak ‘how’
<i>Positive Evaluation</i>			
–	afärin ‘bravo! well done!’	aferin ‘bravo! well done!’	molodets ‘bravo! well done!’
–	amin! ‘amen’	–	–
–	inshalla ‘God willing’	–	–
–	–	–	chudo! ‘cool!’
–	–	–	priko! ‘how amusing!’
–	–	–	tochno ‘precisely, exactly’
–	–	–	voobshche klassno! ‘totally cool!’
<i>Negative Evaluation</i>			
–	–	–	uzhas! ‘how awful!’
–	–	–	zhalko! ‘it’s a shame!’
<i>Epistemics</i>			
móc ‘to be able’	–	–	–
–	bälki ‘probably’	belki ‘probably’	naverno ‘probably’
–	ikhtimal ‘probably’	–	–
–	mögayin ‘probably’	–	–
–	mömkin ‘possible, likely’	–	mozhet ‘maybe, possibly, might’
–	(as ‘absolutely’ in Tatar)	mutalaka ‘without a doubt’	–
<i>Hedges</i>			
–	–	–	chut’ ‘almost’
–	–	–	khot’ ‘at least, only’

Russian/Polish in Karaim	Arabic/Persian in Tatar	Turkish in AMGreek	Russian in Tatar
okolo 'around'	–	–	okolo 'around'
–	–	–	pochti 'almost'
–	–	–	prosto 'simply, just'
–	–	–	ni v koem sluchae 'by no means'
–	–	–	sluchaino 'accidentally'
<i>Evaluations of Degree</i>			
–	fäkat" 'only, just'	–	tol'ko 'only, just'
–	gajäp 'surprisingly'	–	–
–	gayat' 'very'	pek 'very'	–
–	–	chep [intensifier]	–
–	–	daha 'still'	eshche 'more, still'
–	shulkadar 'so, as much as that'	o kadar 'so, as much as that'	–
–	ifrat 'overly'	–	slishkom 'overly'
–	kadär 'as much as'	kadar 'as much as'	–
–	–	beraber 'equal to, uniform with'	–
–	–	tïbkï 'quite like, similar to'	–
–	–	dek 'as far as, until'	–
–	khattä 'even'	–	dazhe 'even'
–	khosusan 'particularly'	–	–
–	tamamän 'completely'	tamamen 'completely'	sovsem 'completely'
–	motalakan 'absolutely'	(as epistemic – see above)	absoluyutno 'absolutely'
<i>Evaluations of Time</i>			
uzhe 'already'	–	–	uzhe 'already'
–	–	–	skoro 'soon'
–	–	–	srochno 'quickly, soon'

Russian/Polish in Karaim	Arabic/Persian in Tatar	Turkish in AMGreek	Russian in Tatar
–	–	birden bire ‘suddenly, at once’	srazu ‘immediately’
–	–	yeniden ‘anew, afresh’	–
<i>Evaluations of Degree</i>			
–	–	eyi ‘yes, very well’	da ‘yes’
–	–	hayir ‘no’	nyet ‘no’
yar ‘each’ (< Persian)	här ‘each, every’	her ‘each, every’	–
–	hich ‘nothing, none’	hich ‘nothing, none’	–
–	hämmä ‘all, each’	–	–
<i>Evaluations of Relationship to Reality</i>			
–	daimi ‘constantly’	–	–
378	gomumän ‘in general’	–	voobshche ‘in general’
–	–	–	obychno ‘usually’
–	–	–	inogda ‘sometimes’
–	–	–	na samom dele ‘in reality, in fact’
–	–	–	poluchaetsia ‘it turns out’
–	–	–	obiazatel’no ‘absolutely, without fail’

Borrowings without discourse-pragmatic function

Russian/Polish in Karaim	Arabic/Persian in Tatar	Turkish in AMGreek	Russian in Tatar
<i>Question words</i>			
—	—	kanda ‘where’	—
—	chandan ‘how many’	—	—
—	cheh ‘what kind of’	—	—
<i>Spatial relators</i>			
kolo ‘by’	—	—	—
s ‘from’	—	—	—
za ‘for’	—	—	—
—	kadar ‘up to’	—	—
—	gan ‘from; to; about’	—	—
—	ila ‘to, up to’	—	—
—	fäuk” ‘on, higher’	—	—
—	äz ‘from’	—	—
—	—	karshî ‘opposite’	—
<i>Pronouns</i>			
—	äna ‘I’	—	—
—	lya shäi ‘nothing’	—	—
—	chiz ‘anything, nothing’	—	—
—	käs ‘anyone, noone’	—	—

Russian/Polish in Karaim	Arabic/Persian in Tatar	Turkish in AMGreek	Russian in Tatar
<i>Spatial relators + Pronouns</i>			
—	gankem ‘from you, about you’	—	—
—	ganna ‘from us’	—	—
—	ganha ‘from her, about her’	—	—
—	ganhe ‘from him, about him’	—	—
—	ganhem ‘from them, about them’	—	—
—	ganhema ‘from them both, about them both’	—	—
—	läh ‘to him, for him’	—	—
<i>Adverbs (and adverbial-type phrases) of manner</i>			
—	mäsrürän ‘happily’	—	—
—	mötäwazigan ‘humbly, quietly’	—	—
—	mötägammidän ‘seriously, assiduously’	—	—
—	möshfikana ‘finely, gently’	—	—
—	—	üze üze ‘floating’	—
—	—	shashkïn shashkïn ‘wondering’	—
—	—	agle agle ‘weeping’	—
—	—	sürünü sürünü ‘dragging’	—
<i>Deontics</i>			
—	zarur ‘necessary’	—	—
—	lyazim (now archaic)	lazim ‘necessary’	—
<i>Other grammatical words</i>			
—	liäjel ‘for, for the sake of’	—	—
—	baraye ‘for the sake of’	—	—