AN ANALYSIS
of
RECENT LOANS
into the
STANDARD UYGHUR LEXICON

What Semantic Distribution & Phonological Interpretation Reveal about Transmission Environment

Nicholas D. Kontovas
University of Chicago, Department of Linguistics
Spring, 2008

0   Abstract

A corpus study conducted on the distribution of Russian and Chinese loanwords in the modern Standard Uyghur lexicon reveals a number of striking inconsistencies with our expectations given the history of population dynamics within the region. Russian loans outnumber Chinese loans by a ratio of almost four-to-one, though census data suggests much more direct population contact between Uyghur speakers and Chinese speakers than between Uyghur speakers and Russian speakers. In terms of semantic distribution, Russian lexemes span a wide range of high-register scientific, technological, and philosophical vocabulary, while Chinese lexemes are more restricted in their scope. This is precisely the opposite of what we might expect to find, considering language modernization policy in the Autonomous Region from the late 1950's to the mid-1980's preferred the direct importation of Chinese lexemes. Furthermore, an examination of the phonology of the borrowings reveals a much greater degree of systematicity among Russian loans than among Chinese loans. The work being presented will attempt to account for these discrepancies by making reference to elements of the socio-linguistic ecology of the environment in which the standard variety of modern Uyghur was developed. In it, we explain how the need for secular education among Uyghurs during the early period of Chinese occupation was met by a number of private schools, opened on the Russian model by wealthy Uyghurs educated in Russia (Bellér-Hahn 2000). These schools are the first in a trend of indigenously motivated efforts towards lexical modernization which favored the adoption of Russian loans, and which – fueled both by overt anti-Chinese sentiment among a portion of the Uyghur population, and by the increasing need to define and preserve a unified Uyghur ethnos – has survived to this day as a preference for the use of Russianisms over Sinicisms in the higher register.

1 Introduction

The modern Uyghur language is spoken by just over ten million people world-wide. Of these, the vast majority – approximately eight million people, as of 1990 (Ethnologue 2007) – live in the present day Xinjiang Uyghur Autonomous Region (henceforth XUAR, or simply Xinjiang)[1] of the People's Republic of China. A map displaying the approximate location of the Uyghur-speaking region in an international context appears below in figure 1, with the relevant region enclosed in a red box.

*fig. 1   Location of the Uyghur-speaking Region* [2]

Throughout its history, the region has also been home to a large number of speakers of other languages. These languages have contributed various amounts of vocabulary to Xinjiang's oasis-dwelling Turkic dialects, now collectively known as Uyghur. While the grammar and basic lexicon of the standard variety of the language (most terms for body parts, simplex verbs, etc.) are undeniably Turkic, up to 75% of the entire lexicon may consist of Perso-Arabic loans (Schwarz 1992:xxii). Older non-Perso-Arabic loans come mostly from Mongolian, with some others from Tibetan and Middle Chinese. The most sizeable recent xenolexical contributions to the Standard Uyghur lexicon have come from Russian and Mandarin Chinese. The following investigation will concern itself only with these latter two sets of borrowings.

---

[1] Where the acronym XUAR is used, it will describe only the region after the establishment of the Autonomous Region in 1949.
[2] Modified from the *CIA World Factbook* political world map.

In order to obtain an idea of the presence of these two languages among the Uyghur-speaking community, it may prove useful to examine the population of probable native Russian- and Chinese-speakers within the region during its recent history. Unfortunately, demographic evidence from before the establishment of the Autonomous Region in 1955 is unreliable, and any estimates from before the invasion of the People's Liberation Army in 1949 are entirely speculative. Census data obtained from the *Xinjiang Statistical Yearbook* displaying Xinjiang's Uyghur-, Chinese- and Russian-speaking populations since the last Chinese takeover appear in figure 2 below.[3]

fig. 2   Uyghur, Chinese and Russian Population of the XUAR 1949-2006 [4]
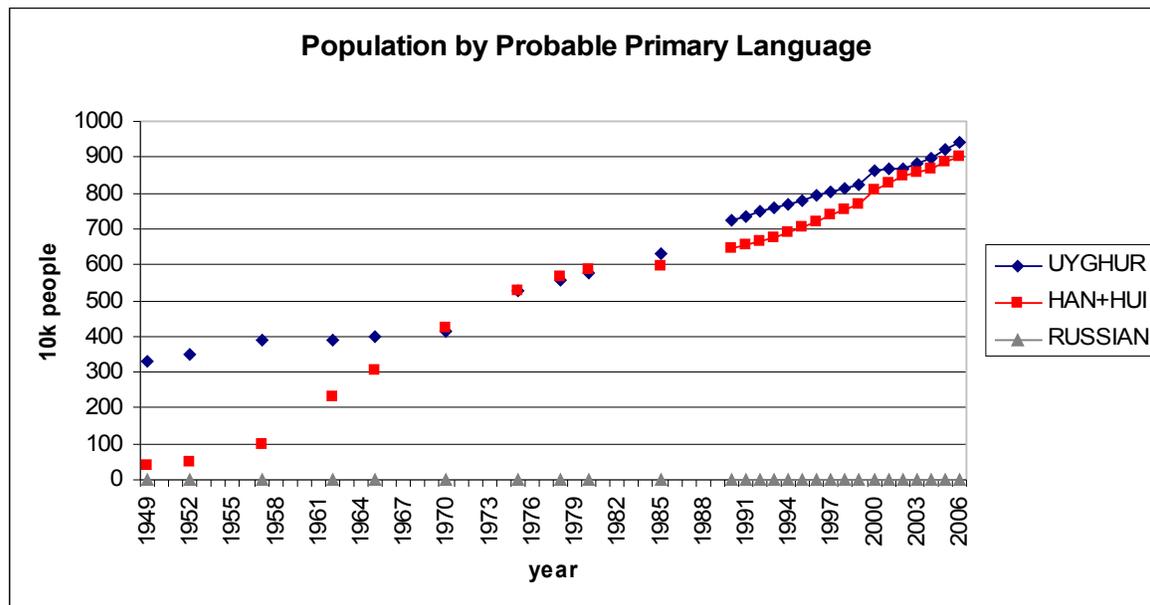
It is perhaps a logical assumption that contact-induced language change requires – and, all other social factors being equal, is proportional to – some degree of contact between speakers of the relevant languages. Thomason and Kaufman (1988) make a point of noting, however, that in situations of borrowing – lexical or otherwise – the contact between the two communities of speakers must be of a sort that it induces a certain degree of bilingualism in the Donor Language (DL) – i.e., the language from which the loanwords (in this case) are being taken – among primary-language speakers of the Recipient Language (RL) – i.e., the language into which the words are being borrowed. (Note that while, in discussing issues of substratum interference, Thomason and Kaufman et al. have occasionally employed the terms "Source Language" and "Target Language" – the latter borrowed from the literature on second-language acquisition – to describe what I refer to here as the Recipient Language and Donor

---

[3] Han Chinese within Xinjiang are not separated into Mandarin-speakers and speakers of other Chinese languages, though the vast majority of Chinese within the region speak a dialect of Mandarin.
[4] Yearly demographic statistics not available until 1990. Statistics for 1949-1978 obtained from the *Xinjiang Statistical Yearbook*, 1991 volume; 1978-2004 from the 2005 volume; 2005 from the 2006 volume; 2006 from the 2007 volume.

Language respectively, I have avoided these terms, as they suggest an overall shift of speakers of the former to the latter.)

That there should exist a correspondence between the degree of bilingualism and the nature of the borrowing from DL to RL is perhaps less immediately obvious. Nevertheless, as far as loanword adaptation is concerned, there is a good deal of evidence to suggest that borrowings that come about as the result of casual contact differ significantly from those that result from intense contact. Weinreich (1953) was the first to note that more significant lexical borrowing tends to induce phonological change in the RL, introducing elements of DL phonotactics, while Kiparsky (1973) suggests that in situations of "extensive bilingualism," RL speakers are more likely to preserve "the distinctness between lexical items of the lending language by means of the phonetic repertoire of the borrowing language" (1973:112). A good example appears in Thomason and Kaufman (1988), where the authors present the following table (in figure 3 below,) noting the increased degree of phonological faithfulness to the individual Russian ancestors of a number of loans in Asiatic Eskimo, corresponding to an increased bilingualism among Asiatic Eskimo speakers in Russian.

fig. 3   Earlier and Later Russian Loans in Asiatic Eskimo (1988:33) [5]

Russian	Earlier Borrowing	Later Borrowing	Gloss
bl'udce	*pljusa*	*bljutca*	saucer
čaj	*saja*	*čaj*	tea
tabak	*tavaka*	*tabak*	tobacco
pačka	*paska(-q)*	*pačka*	bundle

Implicit within these claims is a continuum not unlike the one presented below in figure 4, in which, past some unknown level of bilingualism in the DL among the RL speech community, the increasing number of loans retains more characteristically DL phonotactics. Consequently, this induces phonological change within the RL.

fig. 4   Traditional Model of Loanword Phonology as a Function of Intensity of Contact
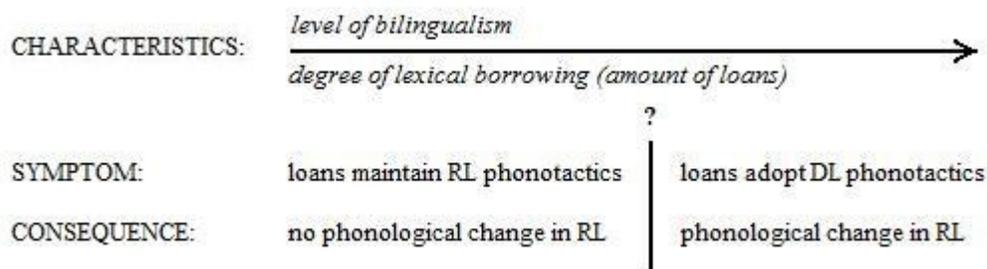
In the pages to follow, I will demonstrate that the distribution of Russian and Chinese loanwords within the Standard Uyghur lexicon shows a startling disparity between the overall level of bilingualism in both languages among the general Uyghur population and the amount of loans from each of the languages into Uyghur.

---

[5] Original data obtained from Menovshchikov (1969:122).

In order to assess the overall distribution of Russian- and Chinese-derived elements within the Standard Uyghur lexicon, and to provide a basis for further investigation, we compiled two corpora: a word-list corpus – designed to give a complete account of all the Russian and Chinese loans available to an educated speaker of standard Xinjiang Uyghur (henceforth referred to as the 'available lexicon,' AL) – and a real-text corpus – designed to give an idea of the Russian and Chinese loans encountered on a daily basis by the average reader within Xinjiang (henceforth the 'real-text lexicon,' RTL.) It should be noted that the goal in compiling these corpora has not been to exclude all bias – bias among language users, in the form of preference for some forms over others, is unavoidable in any situation and will indeed be seen to play an important role in the early composition and maintenance of the Standard Uyghur lexicon – but to exclude any bias which is not inherent in the language as employed by the majority of Uyghurs participating in the every-day communication with defines the boundaries of the acceptable Xinjiang Uyghur lexicon. To that end, care has been taken to ensure that the sources comprising the corpora reflect the language as it appears within the XUAR (excluding the possibility of a pro-Russian bias that might result from the inclusion of Soviet Uyghur texts,) and that the items included in the corpora would be recognized as 'standard' by the average Uyghur speaker (excluding the possibility that heavily Russian- or Chinese-influenced vernaculars, not widely understood among the entire Uyghur-speaking population of Xinjiang, might be over-represented in the corpora.)

In generating the word-list corpus, I have relied heavily on the bilingual Uyghur-English dictionary compiled in 1992 by Henry Schwarz of Western Washington University. In the preface to his dictionary, Schwarz explains how his efforts in compiling the dictionary focused mostly on translating and amending the seminal *Uyghurche-Xenzuche Lughet* ('Uyghur-Chinese Dictionary',) originally published in Ürümchi in 1982. After the initial translation, Schwarz cross-examined the dictionary with a number of sources published in Xinjiang since the establishment of the XUAR, adding well-attested forms which did not appear in the Chinese dictionary and dropping forms for which there was little or no attestation in the literature. The authenticity of the language within the dictionary was further ensured with the help of numerous proofreaders, among whom Abdurop Polat and Litip Toxti, two well-established names in Uyghur linguistics, both based in Ürümchi (Schwarz 1992:vii-viii). The dictionary was carefully examined, and all Russian- and Chinese-derived words contained therein notated, along with the relevant Russian or Chinese ancestor. In an attempt to account for the Schwarz dictionary's being slightly out-of-date, and the possibility that the contents of the dictionary reflect a preference for more Russian loans (on the whole more recognizable to the dictionary's Anglophone audience.) I have further cross-examined the resulting word-list with a number of other Xinjiang-based Uyghur-language dictionaries. The online Yulghun Uyghur-Chinese-English dictionary at http://dict.yulghun.com/ has proved indispensable, and its emphasis on Xinjiang Uyghur is exemplified by its exclusive use of the *Uyghur Erep Yéziqi* (the government-endorsed 'Uyghur Arab Script') in its representation of all Uyghur text. The Uyghur-English dictionary at http://www.uyghurdictionary.org/ – edited in part by Xinjiang University's Waris Abdukerim Janbaz – provides etymologies for any non-Turkic word, and as such has been essential in cross-examining derivations proposed in Schwarz 1992. Other dictionaries employed in the corpus-building process include the 2007 version of the

electronic *Irpan* Uyghur-Chinese-English dictionary, created in Ürümchi and intended for Uyghurs familiar with Chinese learning English, and Qadir Qawuz's *In'gilizche-Uyghurche Lughet: Oqughuchilar üchün* ('English-Uyghur Dictionary for Students',) published by the Xinjiang People's Hygienic Press, also in Ürümchi.

The creation of the real-text corpus is ongoing. At the time of writing it contains words derived from Xinjiang University's *Irpan: In'giliz Tili Gülzari* (a book of English texts with Uyghur translations, designed for Uyghurs learning English) and the example sentences contained within the *In'gilizche-Xenzuche-Uyghurche Disniy Balilar Resimlik Lughiti* ('English-Chinese-Uyghur Disney Children's Dictionary',) published by the Xinjiang People's Press. In the future, radio and television news transcripts may also be included.

After initial compilation, each corpus was analyzed in order to obtain a ratio of individual Russian to Chinese loans (henceforth R:C), expressed below as a number of Russian lexemes for every one Chinese lexeme. Out of an AL of approximately 21,000 total entries, 1,035 were derived from Russian, and 296 from Chinese, yielding an overall R:C of 3.33. In the RTL as it exists thus far (containing approximately 5,000 words,) there appear 53 unique Russian loans, and only 3 unique Chinese loans, yielding an R:C of 17.66. A graphical representation of the R:C values in terms of percent total value for the recent xenolexical contribution in both corpora appears below in figure 3.

fig. 5   Ratio of Russian to Chinese Loans in Each Corpus
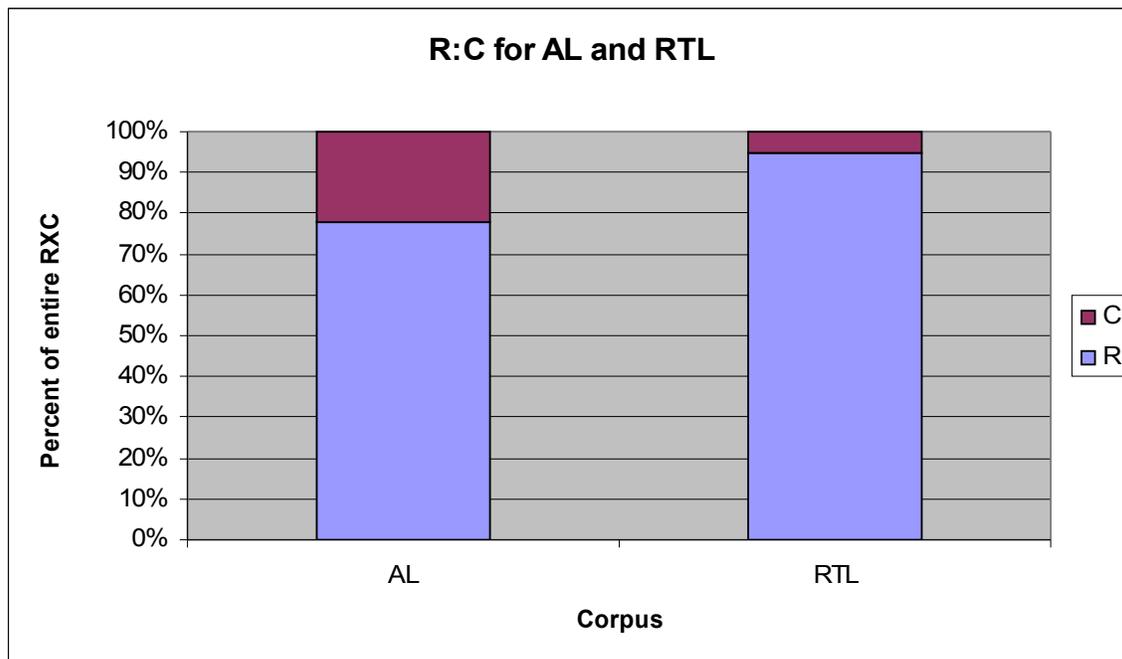
The implication of the above results is that there are considerably fewer Chinese loanwords than Russian loanwords in the Standard Uyghur lexicon. The distribution within the Real-text Lexicon thus far further suggests that, in everyday texts, writers (for whatever reason) make use of the Chinese component considerably less than the Russian.

An analysis of Russian and Chinese loans in modern Standard Uyghur seems, therefore, problematic for any theory of contact-induced language change grounded in the notion that the intensity of contact between populations of speakers is directly correlated with the degree of lexical borrowing. As demonstrated in figure 2 above, the number of Russians in Xinjiang was, and continues to be, very low in comparison to the number of Mandarin-speaking Hui/Tungan and Han. It has likewise constituted only a fraction of the region's combined Uyghur, Russian, and Chinese population. There is furthermore no indication that the population of ethnic Russians wields any significant amount of social capital; they engage largely in agriculture, and very few progress to university-level education (Yu 2004).

The following investigation constitutes a detailed examination of the recent Russian and Chinese contributions to the Standard Uyghur lexicon, based on the assumption that the characteristics of a borrowed lexeme within a language can reveal clues to the nature of the transmission environment surrounding that lexeme. Since, at least in theory, the words contained within the RTL are only a subset of those contained within the AL, it is the latter, and not the former, which will serve as the basis for our analyses. **Section 2** will concern itself with the semantic distribution of words within both Russian- and Chinese-derived elements. **Section 3** will consists of an in-depth investigation into the peculiarities of the phonological interpretation of each element, singling out particular interpretational markers as indicative of certain aspects of the transmission environment for the loans which display them. Given the evidence obtained from both the semantic and phonological analyses, we will then argue in **Section 4** that the modern Standard Uyghur lexicon reflects an evolution grounded in the cultivation of a unified Uyghur identity among the educated elite (especially in the face of a perceived threat from Han Chinese,) consistent with what we know about the sociolinguistic environment of the Uyghur-speaking region.

I will also make a case for the importance of considering strong language attitudes among speakers as a factor in determining the success or failure of attempts at lexical modification. This is especially true in environments of the sort of high-stakes renegotiation of ethno-linguistic identity characteristic of minority language communities. In cases where lexical maintenance is perceived as necessary but attitudinal factors motivate the avoidance of modernization along the model of the dominant language, speakers may – circumstances permitting – intentionally draw upon other sources which, although less readily available, pose less of a threat to their linguistic autonomy. In the Uyghur example, the speech of the influential minority – well-educated, if not exactly fluent, in Russian – became the model for lexical modernization in defiance of the more prevalent, but less linguistically desirable, Chinese. With standardization, the population is progressively more exposed to the model variety, exaggerating the effects of the ideologically preferable DL upon the RL's lexicon. The degree of lexical borrowing and consequent phonological faithfulness of the loans is thus divorced from the degree of overall contact between the RL-speaking population as a whole and the DL, though it crucially does not invalidate the notion that increased bilingualism and increased phonological faithfulness among loans are, at some level, directly correlatable.

2 Semantic Distribution

Each Russian- and Chinese-derived element within the AL was classified according to one of ten categories (with en eleventh 'Miscellaneous' classification for elements not easily fitting into any semantic category.) An explanation of each category appears below in figure 6.

fig. 6 Description of Semantic Categories

Category Name	Description
TRANSPORTATION	types of vehicles; words belonging to the sphere of air, rail, or nautical transportation technology
ART & ARCHITECTURE	styles of art/architecture; trends in artistic/architectural thought; art supplies
SCIENCE	chemical names; modern medicine & medical supplies; advanced scientific / mathematical schools / concepts
PLANTS & ANIMALS	non-culinary plants and animals
SOCIO-POLITICAL	social movements; political offices / positions / bodies / parties
QUOTIDIAN	food; clothing; names of the months; household items
TECH, MANUF, & INFR (technology, manufacturing & infrastructure)	manufacturing tools / equipment; postal terms; mining terminology; non-area-specific technology
LANGUAGE & WRITING	linguistic concepts; typesetting / publishing terms; literary elements (theme, plot, etc.)
TOPO- & ETHNONYMS	Names of places / ethnicities / nationalities
RELIGION & PHILOSOPHY	(non-Islamic) religious items / festivals / concepts; philosophical movements

Even with the possibility of a miscellaneous classification, some ambiguity is expected. Furthermore, even among items which have been classified into one of the ten groups above, some semantic overlap is unavoidable. The categories employed, though at some level arbitrary, were the result of careful consideration of the words within the corpora as a whole, and of the likely environment of initial transmission for each individual loan. An item like *bisey* (< C *bai cai* 'bok choy') for example, though it describes a plant, would be placed into the QUOTIDIAN category, due to its extensive use in cooking.

The distribution of the loans among the eleven semantic categories was then analyzed for both the Chinese and Russian contributions to the AL. The results appear below in figure 7 and figure 8, respectively.

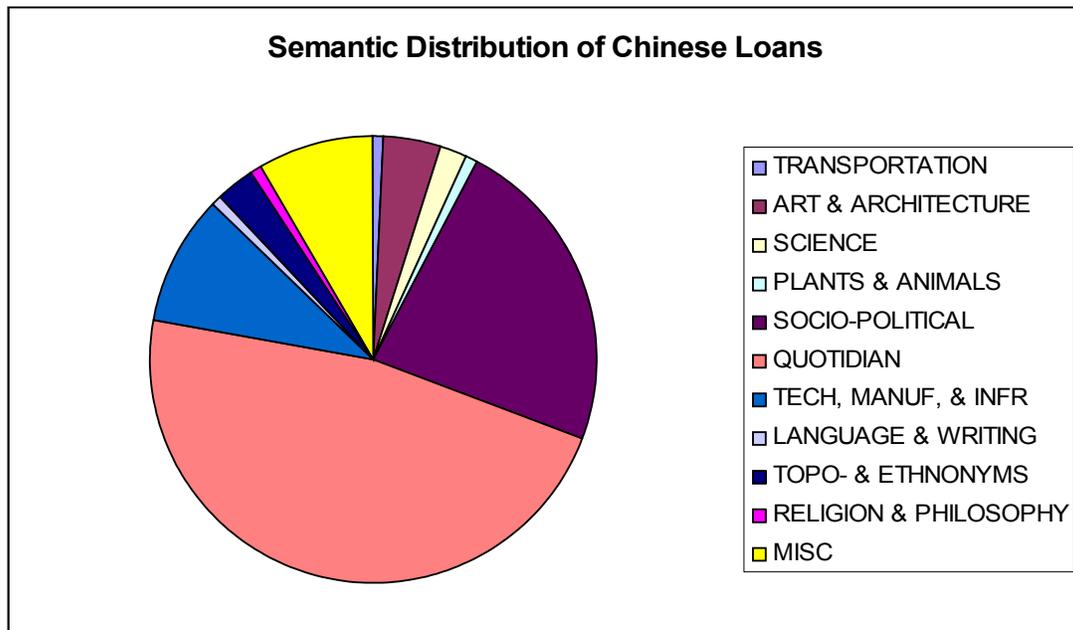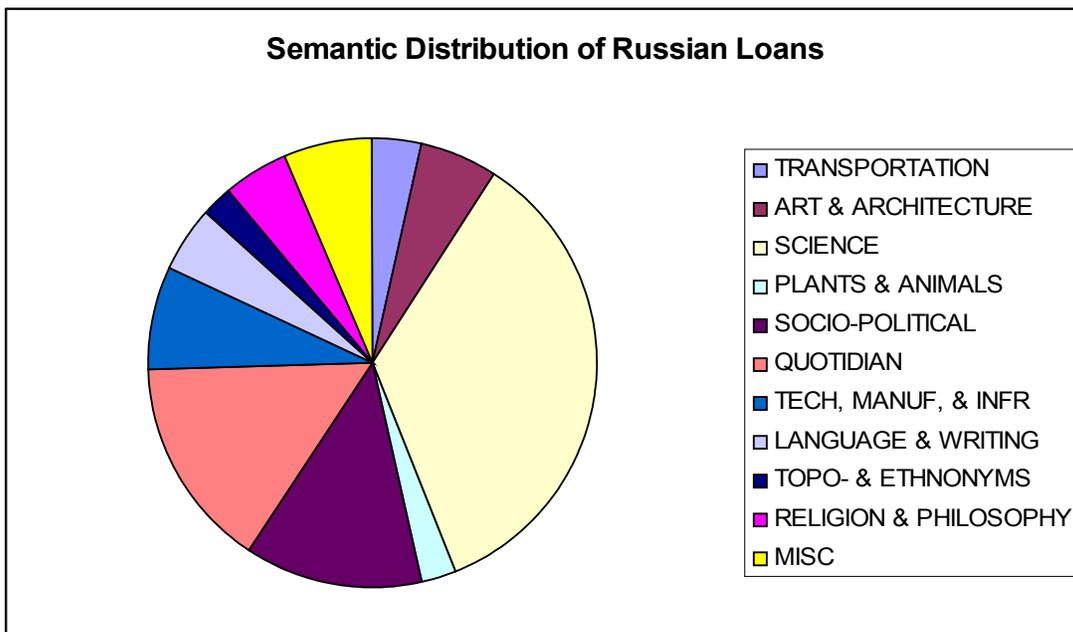*fig. 7   Semantic Distribution of Chinese Loans within the AL*

fig. 8   Semantic Distribution of Russian Loans within the AL

As shown above, the largest portion (34%) of Russian loans in Standard Uyghur belongs to the category of scientific vocabulary, while the largest portion of Chinese loans (47%) belongs to the category of 'everyday' vocabulary. Likewise, religious and philosophical terminology accounts for 5% of the Russian component, while it is completely absent from the Chinese. Such a distribution suggests that Russian loans are

largely confined to the higher registers of the available Standard Uyghur lexicon. The majority of Chinese loans – as might be expect, given the rather prolonged history of contact with between populations of Uyghur and Mandarin speakers – appear in the portion of the lexicon devoted to everyday items.

3   Phonological Interpretation

Much of our attempts to draw parallels between the forms of Uyghur loans and their DL ancestors will rely on the notion of systematic faithfulness to DL phonotactics, well-established within the literature on language contact and exhibited in the chart from Thomason and Kaufman (1988) in figure 3 above.

  Faithfulness, for the purposes of this analysis, has been defined largely according to the similarity of the system of interpretation employed in order to derive a given loan to the systematic interpretational systems currently in place in the former USSR (in the case of Russian) and the PRC (in the case of Chinese.) As demonstrated in figure 10 (following page,) the systems display a largely one-to-one correspondence between DL and RL phonemes,[6] though in the case of Chinese loans, the correspondences are perhaps more easily conceived of in terms of syllable onsets and rhymes.

  The recent xenolexical contribution of 1,331 lexemes contained within the Available Lexicon – of which 1,034 Russian and 297 Chinese – was first analyzed with respect to each lexeme's degree of faithfulness. Any item exhibiting a deviation from the system of normal correspondences was marked, and its interpretational abnormalities labeled. The list of possible interpretational abnormalities appears below in figure 9.

fig. 9   Interpretational Abnormalities[7]

	Label	Meaning	Example	Inconclusive?
Russian	X > Y	abnormal change of segment X to segment Y	/šepke/ 'western-style hat' (< R /šapka/)	(*)
	DE_PAL	pre-vocalic palatalization not rendered	/pulimot/ 'machine gun' (< R /pulim'ot/)	
	TRUNC	whole lexeme is severely truncated	/éléktir/ 'electricity' (< R /el'ektirčestvo/)	
	EP_i	inserts epenthetic or prepenthetic /i/	/marksizim/ 'Marxism' (< R /marksizm/)	*
	X > Ø	abnormal deletion of segment X	/konyek/ 'cognac' (< R /konyjak/)	
	ADD_FIN_a	adds final /a/ to the Uyghur reflex	/nérwa/ 'nerve' (< R /n'erv/)	*
	CLUST_SIMP	consonant cluster simplified	/boğaltir/ 'accountant' (< R / buxgaltir/)	
	CONF	confounded etymology	/üstel/ 'table' (< U /üst/ 'top' + R /stol/)	
	DE_GEM	degemination	/disértatsiye/ 'dissertation' (< R /diss'ertacija/)	*
	FIN_DE_VOICE	final consonant devoicing	/mikrop/ 'microbe' (< R /mikrob/)	*
	VOICE	voicing of an originally unvoiced segment	/kardon/ 'cardboard' (< R /karton/)	
	RED_o	unstressed /o/ rendered as /a/	/marožni/ 'icecream' (< R /moroženoje/)	
Chinese	X > Y	abnormal change of segment X to segment Y	/jado/ 'hand scythe' (< C /zha dao/)	(*)
	DEL_i	deletes Chinese /i/ at the beginning of a rime	/menjaŋ/ 'a type of noodle soup' (< C /mian zhang/)	(*)
	X > Ø	abnormal deletion of segment X	/osuŋ/ 'a type of lettuce' (< C /wo sun/)	
	DEL_u	deletes Chinese /u/ at the beginning of a rime	/chenze/ 'rafter, beam' (< C /chuan zi/)	

---

[6] With a few notable exceptions. In the case of S /-iye/ (< R –ija,) the line indicating a systematic correspondence between ancestor and reflex is perhaps more accurately drawn at the morphemic level. Furthermore, Russian palatalization is treated as an aspect of the vowel and not of the preceding consonant. In Chinese loans, C /zi/ appears alternatively as /za/ or /ze/, harmonizing with the quality of the vowel that precedes it.

[7] Interpretational abnormalities which amount to nothing more than orthographic variants with the same pronunciation in Uyghur have not been considered.

*fig. 10 Normal Sound Correspondences for Russian & Chinese Borrowings in Uyghur*

Russian	Uyghur
/a/	/a/
/'a/	/ja/
/b/	/b/
/c/	/ts/
/č/	/č/
/d/	/d/
/e/	/é/
/'e/	/je/
/f/	/f/ (/p/)
/g/	/g/
/x/	/x/
/i/	/i/
/j/	/y/
/ž/	/ž/
/k/	/k/
/l/	/l/
/m/	/m/
/n/	/n/
/o/	/o/
/'o/	/jo/
/p/	/p/
/r/	/r/
/s/	/s/
/š/	/š/
/t/	/t/
/u/	/u/
/'u/	/ju/
/v/	/w/
/y/	/i/
/z/	/z/

Chinese	Uyghur
b-	/b/
c-	/s/
ch-	/č/
d-	/d/
f-	/f/ (/p/)
g-	/g/
h-	/x/
j-	/j/
k-	/k/
l-	/l/
m-	/m/
n-	/n/
p-	/p/
q-	/č/
r-	/r/
s-	/s/
sh-	/š/
t-	/t/
w-	/w/
x-	/š/
y-	/y/
z-	/z/
zh-	/j/
-a	/a/
-ai	/ey/
-ao	/aw/
-an	/en/
-ang	/aŋ/
-e	/é/
-ei	/éy/
-en	/én/
-eng	/éŋ/
-i	/i/
-ia	/ya/
-ian	/yen/
-iang	/yaŋ/
-ie	/yé/
-iu	/yu/
-iong	/yoŋ/
-o	/o/
-ong	/oŋ/
-ou	/ow/
-u	/u/
-uan	/wan/
-uang	/waŋ/
-uo	/wo/
-ü	/ü/
-üan	/üen/
-üe	/üé/

Interpretational abnormalities in figure 9 (above) with an asterisk in the rightmost box are inconclusive when determining the faithfulness of the original form of the loans which display them. EP_i is inconclusive, as it is optionally inserted in order to break up

consonant clusters in virtually any Uyghur word,[8] and therefore not necessarily a product of imperfect transmission. ADD_FIN_a represents a similar method of dealing with word-final clusters. DE_GEM is also inconclusive, as wherever Schwarz has put a degeminated form, other dictionaries can be found with a geminate variant. FIN_DE_VOICE is inconclusive, due to a native Uyghur rule which devoices all final consonants. Asterisks contained within parentheses represent abnormalities which are only occasionally inconclusive. X > Y in Russian loans, for example, is conclusive for, say, X = /g/ and Y = /j/, but inconclusive for X = /e/ and X = /i/, since the normal corresponding vowel for Russian /e/ is Uyghur /é/, which alternates freely with /i/ in the speech of most Uyghurs. The same holds true in Chinese loans when X = any of the set of Chinese rimes containing vowels corresponding to Uyghur /é/. The Chinese-loan-specific abnormality DEL_i is interesting, in that is it inconclusive only for items in which the relevant syllable in the Chinese ancestor has as its initial consonant an alveopalatal consonant (i.e., /j, q, x/.) This is presumably because palatalization is not permissible in some dialects of Uyghur on the alveolar equivalents to Chinese alveopalatals; whatever the case, any form which in Schwarz displays DEL_i does not in at least one other dictionary, and vice-versa. In general, despite the large number of Russian loans and a high level of bilingualism in Chinese, Uyghur phonotactics do not seem to have been greatly effected.

Based on the interpretation abnormalities exhibited, each Russian- and Chinese-derived item in the AL was labeled as either deviant (D-form) or faithful (F-form.) We then compared the ratio of D- to F-forms for both the Russian and Chinese elements of the derived recent xenolexical contribution. The results appear in figure 11 below.

fig. 11 D-forms to F-forms for both Russian and Chinese Loans

---
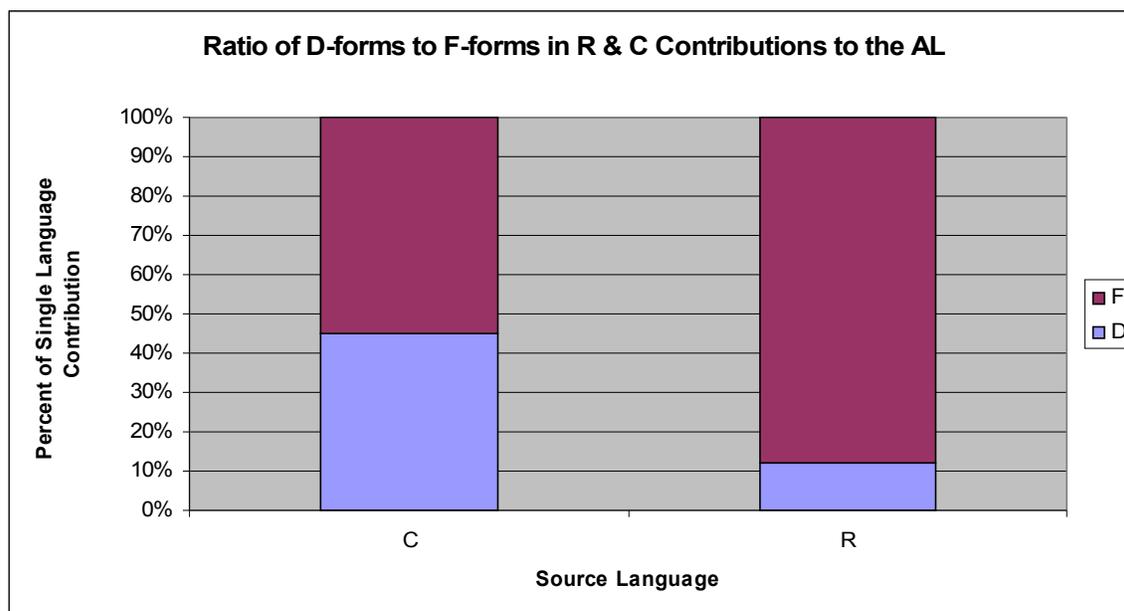
[8] Ex., /ders/ 'class' (< Arabic /dars/) appears optionally as /deris/, depending upon the speaker's personal preference, as well as the level of formality of the situation.

As displayed in the chart above, D-forms account for approximately 45% of the entire Chinese contribution to the Available Lexicon, but only 12% of the entire Russian contribution, suggesting that Russian forms display an overall greater degree of phonological faithfulness to their DL ancestors than Chinese forms. This is again, as suggested by the literature outline in **Section 1**, not surprising given the relatively large amount of Russian borrowings in Uyghur. The findings do, however, suggest a high degree of bilingualism surrounding the transmission of the remaining 88% of the Russian-derived element of the lexicon – problematic, considering the very low degree of bilingualism in Russian among the general Uyghur population.

In addition to the argument based on systematic faithfulness, we may also make arguments based on individual aspects of interpretation and their likely implications in terms of the transmission environment of the lexemes which display them. A large number of Chinese F-forms, for example, are rendered orthographically exactly as if they had been transliterated grapheme-for-grapheme from romanized spellings (in Hanyu Pinyin) of their Chinese ancestors, despite the fact that their pronunciation in Uyghur could be – and, in some spelling variants, frequently is – rendered more accurately. For example, the first syllable of the word /šawjaŋ/ 'major general' (< C /shao jiang/) – pronounced [ʃaʷʤaŋ] – is alternatively spelled *shao*, *shau*, and *shaw*. This is not surprising, given that the majority of Uyghurs have had at least some education in Chinese; what is surprising is that Uyghurs with a knowledge of the Chinese romanization would persist in choosing a nativized spelling of Chinese-derived lexemes.

A number of Russian F-forms display the interpretational feature PRES_o, which preserves the unstressed Russian /o/ (pronounced [a]) as Uyghur /o/, suggesting a knowledge of the original Russian orthography. The feature is in contrast to the RED_o of some Russian D-forms, which reduces the unstressed /o/ to /a/, more closely resembling the DL pronunciation. Not coincidentally, the former appears mostly in scientific and technological vocabulary – such as /xolodélnik/ 'freezer, icebox' (< R /xolod'el'nik/) – while the latter is exhibited by words describing articles of clothing or foods, ex. /marožni/ 'icecream' (< R /moroženoje/.)

The less artificial, more traditional nature of the contact responsible for the sizeable number of Chinese D-forms is attested by a number of features. NO_JS – no *juan she*, or retroflexion – denotes a lexeme which shows signs of having been adopted from a non-standard variety of Mandarin, in which retroflex consonants are pronounced alveolarly. The feature ADD_ZI marks lexemes which have been interpreted with a final /-za/ or /-ze/ in Uyghur (< C /zi/,) though in Standard Mandarin, the corresponding lexeme contains no /zi/. According to Yakup (2005), /zi/ acts as a productive diminutivizer in the Hui/Tungan dialects historically spoken in and around Turpan (a city in central Xinjiang.) As an interesting aside, a number of the non-systematic correspondences among Chinese D-forms match the pattern of sound correspondences historically followed by Mandarin loans into the Turpan dialect of Uyghur, as outlined by Yakup (2005:175). The prolonged contact between Tungans and Uyghurs in this region, combined with subsequent dialect leveling during the period of standardization, could explain the origin of the deviations displayed by a large portion of Chinese D-forms. It is probable that a parallel situation with Russian among dialects in northern Xinjiang is the ultimate source of the Russian D-forms, however no extensive work has been done on

Uyghur dialects in close, prolonged contact with Xinjiang's autochthonous Russian-speaking population.

On the whole, our analysis of the phonological interpretation of Russian and Chinese loans in Standard Uyghur reveals a striking degree of faithfulness among the Russian-derived component to original DL phonology, suggesting that the group responsible for transmission maintained at least some degree of fluency in Russian. Particular aspects of the interpretational system employed in deriving the Russian borrowings further suggest that the original transmitters of the F-forms had some knowledge of Russian orthography. Chinese loans, by contrast, are shown to be largely the result of direct contact between semi-bilingual Uyghur- and Chinese-speaking populations, whereas the nuances of the interpretational system employed in deriving the majority of Chinese D-forms suggests a possible origin in the Uyghur dialect of Turpan-Qumul.

4   Contextualizing the Standard Uyghur Lexicon

Armed with the results obtained from the preceding analyses, we can now begin to search for a period (or periods) of time which, according to extra-linguistic evidence, might correspond to the transmission environment of a sizeable part of the Russian composition to the Standard Uyghur lexicon, i.e. the Russian F-forms.

Evidence from our investigations of the phonological interpretation and semantic distribution of Russian loans in Standard Uyghur suggest a highly bilingual, highly literate, high-register (likely scientific) transmission environment. A quick analysis of a number of texts written before the 1920s (Gunnar Jarring's turn of the century *Literary Texts from Kashgar*) clearly shows that the Turkic varieties of Xinjiang, at least in their literary form, contained few, if any, Russian loans. Moreover, there is nothing to suggest that the inhabitants of the area engaged in the sorts of activities which might warrant the need for an extensive scientific vocabulary. Contact with the smaller ethnic Russian population in Northern Xinjiang might have been sufficient to facilitate the transfer of the relatively small number of Russian D-forms, but cannot account for any higher-register borrowings.

The first evidence of the systematic introduction of concepts requiring a modern scientific and technological vocabulary begins with the founding of with first *penni* ('scientific') schools in Kashgar around the end of the nineteenth century. These schools – opened in response to the need for secular education within the region, not met by the existing Qing dynasty administration – were inspired in large part by contemporary Russian models, and were often opened by wealthy Uyghurs who had studied abroad in the Russian Empire and later USSR. Many also included Russian language as a subject of instruction (Bellér-Hahn 2002:57). The existence of these schools continued well into the 1930s, gaining in influence through different circumstances in Northern and Southern Xinjiang.

In the south, the combined efforts of the *Association for Uyghur Cultural Enlightenment* and the *Kashgar County Education Office* led to the founding of a number of schools structured along the new, Soviet-inspired model. The schools were largely staffed by former students of the *Kashgar Teacher's Training College*, the majority of

whose textbooks were printed within the Soviet Union. This period saw an increase in regional school attendance rates among wealthy Uyghur children to about 6.9% (2002:63-64) the highest before the advent of the Soviet-backed Turkic Islamic Republic of East Turkestan in 1933. In Northern Xinjiang, Sovietization of the regional educational system gained strength under the regime of Sheng Shicai, a local warlord with close ties to the USSR. In 1942, Sheng lost favor with the Soviets, and the USSR seized the opportunity to back indigenous efforts to establish a Second East Turkistan Republic. The Republic was officially established in 1944 and persisted until 1949, when the People's Liberation Army gained undisputed control over most of the region. The Chinese Communist Party would continue to employ Soviet educational planners until well into the 1950s however, and by all accounts education during this period showed no significant changes.

In light of the fact that, throughout the past fifty years of Xinjiang's history, there has been no significant increase in the Russian-speaking population and considerable increase in both the size and prestige of the Chinese-speaking population, the factors contributing to the maintenance of such a sizeable Russian component to the Standard Uyghur lexicon bear explanation.

One likely reason for the relative paucity of Chinese intrusions into the higher register of the Uyghur lexicon is the lack of enrolment in institutional education among the Uyghur-speaking population during the period of greatest Chinese involvement in minority language planning. Until recently, periods of Chinese dominance in the area of linguistic engineering have coincided with a negative attitude towards the educational avenues by which elements of the standardization program might be diffused among the Uyghur-speaking population. Bellér-Hahn (2000) notes that in the 1950's, just as we begin to see an influx of Chinese-speakers to the region and a decrease in the influence of the Soviet system upon education in Xinjiang, a decline in the overall quality of education prompts significant withdrawals from – and a certain degree of distrust towards – institutionalized schooling. The period of forced Sinicization of the lexicon during the 1950s[9] likewise saw low school attendance among Uyghur children, while the complete dismantling of the PRC's educational system during the years of the Cultural Revolution (1966-76) further limited the sort of contact between Uyghur- and Chinese-speakers which might have facilitated the transfer of higher-register vocabulary.

In 1984, not long after the region's educational system had begun to recover from the crippling effects of the Cultural Revolution, the central government of the PRC in Beijing passed the *Revised Law on Regional Autonomy*. The law reaffirms the right to linguistic self-determination among ethnic minorities within the Autonomous Regions. This has in theory both granted increased freedom in terms of lexical modernization to Uyghur intellectuals (primarily those comprising the Xinjiang Language and Script Committee,) and increased support for the publication and dissemination of educational and entertainment media in Uyghur, facilitating the dissemination of the significantly Russified standard. Dwyer (2005) suggests, however, that the effects of the law are undermined somewhat by an "unequal division of power and resources,"[10] and it is unclear to what extent we can ascribe the continued success of Russian-derived lexemes in Standard Uyghur to its passing.

---

[9] For more on past policies of lexical Sinicization in the XUAR, see Dwyer 2005, pg. 27.
[10] Dwyer 2005, pg. 50

More important than a lack of access to Chinese-language education however, is the continued perception among Uyghurs of the threat posed by an increasing Chinese presence to their ethnolinguistic autonomy. In this way, the issue of the composition of the standard variety of the Uyghur language is only part of a larger trend over the past century and a half towards greater definition of a unified Uyghur ethnos. Ruddleson (1998) explains how certain strata of eastern Turkestan's Turkic-speaking population – faced with the threat of Han Chinese encroachment – have promoted the creation of a single Uyghur identity out of the region's formerly distinct oasis-dwelling populations. The Russian contribution to the Standard Uyghur lexicon – its introduction having coincided with the beginning of the period of Uyghur ethnogenesis, and itself having occupied a previously empty semantic niche – became, in the minds of Uyghur speakers, an inextricable part of their socio-linguistic identity. In accordance with the founder effect principle (Mufwene 1996), such a part would, by simple virtue of the fact that it was a previously established element of the lexicon, be difficult to dislodge, especially under conditions in which speakers stand to gain something by continuing to employ it. In this case, yielding to linguistic pressures to adopt Chinese lexemes may very well constitute a sacrifice of the speaker's ethnic identity. To be sure, anecdotal evidence (among which, disparaging comments made to those who use Chinese words in Uyghur-language internet forums) supports the theory that the use of Chinese lexemes in Uyghur is frowned-upon.[11]

A number of outside influences may also contribute resources which have enabled the maintenance of the Russian element of the Standard Uyghur lexicon. Kamalov[12] suggests that Xinjiang Uyghur intellectuals maintain some links to the sizable Uyghur communities of the post-Soviet Central Asian republics, largely bilingual in Russian. The nature of the relationship between the two populations is not clear, however, and further investigation is required to determine if contact with Soviet Uyghurs is a significant factor in sustaining the Russified portions of the lexicon within Xinjiang. Contact with Central Asian Turkic language themselves may also have some effect on maintaining Russian loans in Uyghur. Paradoxical though it may seem, a desire for closer lexical proximity to other Turkic languages may also play a role. Central Asian Turkic languages – such as Uzbek and Kazakh – make considerable use of Russian loans, despite the efforts of language planners in their respective states to do otherwise. Again, it is not entirely certain what percentage or socio-economic cross-section of the Xinjiang Uyghur population is capable of reading – or indeed, accessing – language materials from the Central Asian states, and therefore not altogether clear that this phenomenon constitutes a significant method of maintenance for the Russian element of the lexicon.

Much of the Russian component of the Standard Uyghur lexicon could also be considered "international" vocabulary, and as such is shared by a number of other languages. Increased contact with languages like English may provide another way of maintaining at least a portion of Russian-derived lexemes. The presence of morphological re-analysis (RE_AN) among a small group of loans may suggest a conscious maintenance of the Russian element to the Standard Uyghur lexicon, though along the model of other

---

[11] Note that, while the increased bilingualism in Chinese among Uyghurs within the XUAR is likely to lead to the increased ability to recognize Chinese elements of the Uyghur lexicon, there has been no research to date on the ability of Xinjiang Uyghurs to recognize Russian loanwords as such.
[12] date unknown

Western languages; for example, the Russian noun /komissija/ "commission" appears in Standard Uyghur both as /komissiye/ and as /komissiyon/, the latter perhaps by analogy with English.

5   Conclusion

While there has never been widespread bilingualism in Russian among Uyghurs within Xinjiang, Uyghur intellectuals and early educational planners, with proficiency in Russian and ties to the USSR, have left a profound impact on the higher registers of the Standard Uyghur lexicon. Evidence further shows that, by the time education and bilingualism in Mandarin began to increase, language attitudes already preferred the maintenance of the existing Russian component to the adoption of newer Chinese loans in most disciplines.

These findings emphasize the importance of a nuanced examination of the population structure surrounding the environment in which a given language-contact phenomenon continues. They also remind of the necessity for periodization in such examinations, and of the overall sensitivity of language evolution to the spacio-temporally bounded aspects of its environment. More specifically, they demonstrate the ability to determine features of the transmission environment of a given loan based on aspects of the loan itself, in the context of a larger xenolexical element.

Further research should focus on expanding the Real-text Lexicon (as described in Section 1) to include more resources representative of the standard language as experienced by speakers on a daily basis, and on exploring the avenues by which the Russian component to the lexicon are maintained.

BIBLIOGRAPHY

Bellér-Hahn, Ildikó. 2000. *The Written and the Spoken: Literacy and oral transmission among the Uyghur*. Berlin, Das Arabische Buch.

Duval, Jean Rahman & W. A. Janbaz. 2006. The Uyghur-English Dictionary Project. Online at http://www.uyghurdictionary.org/.

Dwyer, Arienne. 2005. *The Xinjiang Conflict: Uyghur Identity, Language Policy, and Political Discourse*. Washington, East-West Center.

Gordon, Raymond G., Jr. (ed.). 2005. Ethnologue: Languages of the World, Fifteenth edition. Dallas, Tex.: SIL International. Online at: http://www.ethnologue.com/

Jarring, Gunnar. 1989. *Literary Texts from Kashgar, Edited and Translated with Notes and Glossary*. Lund: Kungl. Humanistika Vetenskapssamfundet.

Jingna'erjiang, Tuohetahong (ed.). 2002. *In'gilizche-Xenzuche-Uyghurche Disniy Resimlik Balilar Lughiti* ['English-Chinese-Uyghur Disney Children's Picture Dictionary']. Ürümchi, Xinjiang People's Press.

Kamalov, Ablet. "Uyghur Studies in Central Asia: A Historical Review". Accessed online at http://www2.toyo-bunko.or.jp/zenbun/NART1/ao_01.pdf

Kiparsky, Paul. 1973. Phonological representations. In Osamu Fujimura (ed.,) *Three dimensions of linguistic theory*. Tokyo: TEC Co., Ltd., 1-136.

Menovshchikov, G. A. 1968. Aleutskij jazyk. *Jazyk narodov SSSR*, vol. 5: *Mongol'skie, Tunguso-Man'chzhurskie i Paleoaziatskie jazyki*. Leningrad, Nauka, 368-406.

Mohammed, Elham (ed.). 2004. *Irpan: In'giliz Tili Gülzari* ['Irpan: English Garden']. Ürümchi, Xinjiang University Press.

Mufwene, Salikoko S. 1996. "The Founder Principle in Creole Genesis". Diachronia 13:1, 83 ff.

Qawuz, Qadir. *In'giliche-Uyghurche Lughet: Oqughuchilar üchün* ['English-Uyghur Dictionary for Students']. Ürümchi, Xinjiang People's Hygienic Press.

Rudelson, Justin J. 1997. *Oasis Identities: Uyghur Nationalism along China's Silk Road*. New York, Columbia University Press.

Schwarz, Henry. 1992. *An Uyghur-English Dictionary*. Bellingham, Center for East Asian Studies, Western Washington University.

Thomason, Sarah Grey & T. Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkley, University of California Press.

Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems*. Reprint: Mouton de Gruyter The Hague, 1963.

Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Malden, Blackwell Publishing Ltd.

Yakup, Abdurashid. 2005. *The Turfan Dialect of Uyghur*. Wiesbaden, Otto Harrassowitz GmbH & Co. KG.

Yu Jianzhong (ed.). 2004. *Eluosizu: Xinjiang Tachengshi Ergongzhen Neimengguer Gunashi Shishuxiang Diaocha*. Kunming, Yunnan University Press.

Yulghun Team, The. 2008. Uyghurche-In'gilizche-Xenzuche Köp Yönilishlik Lughet ['Uyghur-English-Chinese Dictionary of Often-Used Words']. Online at: http://dict.yulghun.com.