

# Discourse pragmatics as a means of contact-induced change

## Suzanne Wertheim

### 1. Introduction

How do the social, cultural, and political backgrounds of minority-language speakers affect the way that they speak? And how does language as a social practice influence language as a system, such that, for example, a minority language becomes more like the dominant language with which it is in contact? This paper uses data from young urban ethnic Tatars, all Russian-Tatar bilinguals, to address these questions and to demonstrate that the code-mixing of discourse-pragmatic words is a previously unrecognized pathway of language influence and contact-induced change.

The ethnographic and linguistic data presented here were gathered during a year of participant-observation fieldwork in Tatarstan, which is an autonomous republic in the Russian Federation. Tatar is a Turkic language found in Tatarstan and elsewhere in the former Soviet Union: currently there are approximately 1 million speakers of Tatar in Tatarstan – one quarter of the Republic’s total population – but like most minority languages in the Russian Federation, Tatar is a “contracting” language undergoing multi-generational language shift, and the number of speakers decreases each year despite “promotive” language policies at the republic level. The speakers studied were bilingual Tatar youth (18-30) residing in Kazan, the capital of Tatarstan. They were all members of a youth organization that I call the “Tatar Social Club,” and either nationalistically oriented or actively engaged in Tatar cultural activities.

In order to remove questions of linguistic competence, the data presented in this paper were taken only from Tatar-dominant speakers or balanced bilinguals. Additionally, all speakers quoted here have demonstrated a markedly pro-Tatar sociolinguistic stance: some choose ideology over communication and periodically refuse to switch to Russian in order to accommodate non-speakers of Tatar, and several have explicitly stated that they “never use Russian unless it is impossible to avoid,” thus violating current sociolinguistic conventions of Tatar use.

Even so, in certain situations when these young bilingual Tatars think they are speaking “pure” Tatar, they are actually using certain Russian words. These Russian words come from different grammatical classes, but they have something in common: they are all used to structure or comment on discourse, they are discourse-pragmatic words. Neither speakers nor listeners appear to be to be aware of these Russian discourse-pragmatic words in otherwise Tatar speech, and they can even sneak through in the most highly performed Tatar style. These Russian words retain their Russian grammatical requirements, and when these requirements are not parallel with those of their Tatar equivalents, the Tatar speech produced is more like Russian. The discourse produced can be interpreted as a “composite” language, not entirely Tatar and not entirely Russian. In Section 7 of this paper, I will show that this discourse-pragmatic code-mixing, and the morpho-syntactic change that it entails, is far from an isolated case.

### 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 paper.

## 2.1 Code-switching vs. code-mixing

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, p. 60). 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).

## 2.2 Discourse-pragmatic words

Studies of language contact usually divide the lexicon into two categories: content words and system words and morphemes. Based on my findings for Tatar, I am proposing that in fact a three-way division of the lexicon is necessary in order to understand contact-induced change: (1) content words; (2) system words and morphemes; (3) discourse-pragmatic words. The code-mixed Russian words found in the speech and writing of the young Tatars described here include particles, adverbs, conjunctions, and question words, but can be grouped into the single functional class of discourse-pragmatic words.

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 otherwise “pure” Tatar include, but are not limited to, discourse markers.

The set of discourse-pragmatic words is usually defined by linguists using heuristics and other linguistic criteria. However, in this paper I am presenting a natural class of discourse-pragmatic words produced by speakers themselves. When speaking in the style in which this code-mixing occurs, Tatar use a negligible number of Russian nouns and verbs, and almost always to fill lexical gaps. When these content words are removed from consideration, all of the remaining Russian words code-mixed by these speakers in otherwise purely Tatar discourse can be seen to have discourse-pragmatic functionality. No single existing definition of discourse-pragmatic words accounts for all of the code-mixed words found in my corpus, but they are all accounted for by at least one theory of discourse or pragmatics.

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)
- conceptual meaning vs. procedural meaning (Sperber & Wilson, 1986; Blakemore, 1987)
- representational function vs. computational function (Rouchota, 1998)
- host syntactic plane vs. disjunct syntactic plane (Espinal, 1991)
- propositional language vs. ‘everything else’ (Fraser, 1996)
- denotative vs. metalinguistic language (Maschler, 1994)

The Russian discourse-pragmatic words code-mixed into Tatar are:

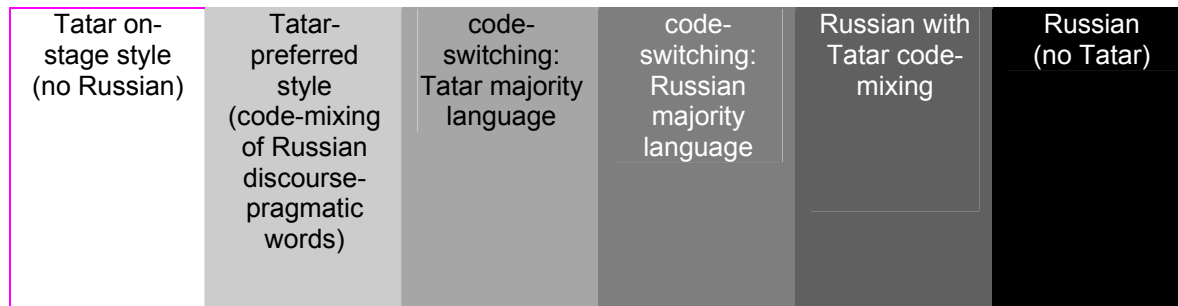
- metalinguistic rather than denotative, but not always procedural in function

- propositional, procedural, or both
- include metatextual words (cf. Traugott’s 1982 models of levels of language)
- not limited to prototypical discourse markers.

I have chosen to use the term “discourse-pragmatic” over many of the other available terms as it is neutral with regard to focus on speaker production vs. hearer reception, and is also neutral with regard to various cognitive theories of speech processing.

### 3. Tatar style shifting

Young urban Tatar bilinguals have a continuum of performance based on language mixing, with pure Russian on one end, pure Tatar on the other end, and a variety of mixed styles in between, as can be seen in the following figure:



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 here. 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, albeit types that are clearly related.

The two most important styles for the analysis presented here are at the Tatar end of the cline and are the pure Tatar style, which I call “Tatar on-stage” style, and the style in which this code-mixing of discourse-pragmatic words takes place, which I call “Tatar preferred” style. 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 found in informal register in private conversations. This is the style with the highest level of self-consciousness and verbal filtering, and is the style used most prominently in the construction of a cultural and ethnic Tatar identity. In Tatar on-stage style, speakers will de-Russify their Tatar to the best of their ability, filtering out as many Russian lexical items as possible. 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. It is not used in formal registers, and the the level of language awareness and verbal hygiene is lower than in the “on-stage” style.

#### 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 neither code-switching nor code-mixing took place (Wertheim 2003a). In order to get the widest range of data and examples of code-mixed Russian words, I used four methods of data collection:

- recordings of conversations (where I was both present and absent)
- field notes (from participant observation where I was either one of the speakers in the conversation or an auditor/“ratified listener”)
- e-mail
- the bulletin board on a Tatar youth “get acquainted” website. All postings used for data were written only by speakers well-known to language investigator, from whom spoken data had already been collected. In addition, the majority of bulletin board posters on this website are members of the “Tatar Social Club.”

#### 5. Thoughts on “pure” Tatar

Style shifting is set within the backdrop of a ubiquitous “discourse of purity” that is found not only among these young Tatarphones, but is also 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. “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 style, code-mixed discourse-pragmatic words slip through (as if they are somehow “flying under the radar”). 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 (although rarely), in the highly performed “pure” Tatar 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 use Tatar in a variety of functional domains, including in professional and educational spheres. In example [1] below, we see the response of one Tatar when asked the question, “What do you think speaking ‘pure’ Tatar means?”

[1] One response to the question “What is pure Tatar?”

Saf tatarcha – ul rus süz-ler-e-n kïstïr-mïycha  
pure Tatar – it Russian word-PL-POSS-ACC insert- NEG.PRES.GER

söyläshü, tïnglau-chï-lar-ga **priyatno** bul-ïrga tiesh tïngla-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.’<sup>1</sup>

This code-mixing of the evaluative Russian adverb *priyatno* in this definition of pure Tatar clearly demonstrates its lack of perceptual salience.

## 6. Examples of code-mixing in the speech of young Tatar bilinguals

When speaking or writing in Tatar-preferred style, young Tatars code-mix a negligible number of nouns and verbs (n = 4 in my corpus); once these words are removed from consideration, it can be seen that all of the remaining Russian words code-mixed in otherwise Tatar discourse have discourse-pragmatic function, and polysemous Russian words that have been grammaticalized will only be used in their discourse-pragmatic functions. These discourse-pragmatic words can be separated into three major groupings: (1) markers of discourse structure and force; (2) interactional performatives; (3) evaluatives.

### 6.1 Markers of discourse structure and force

#### 6.1.1 Subordinating discourse markers

The following words are used as discourse-structuring subordinating conjunctions:

- *chto* ‘that’ (subordinating conjunction)
- *chtob(y)* ‘in order to, that’ (subordinating conjunction)
- *potomu chto* ‘because’ (subordinating conjunction)

In example [2] below, we see an example of an utterance with a code-mixed *chto* used as the subordinator – the subordinate clause is underlined.

[2] Code-mixed *chto*

Min shat sin-ga, **chto** sin taza-sau häm kör küngel-le.  
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.’

<sup>1</sup> The following conventions are used here for the presentation of Tatar data: Tatar words are in plain typeface; Russian words, along with their translations, are presented in **boldface**; conventional Russian borrowings are presented as Tatar, in plain typeface.

The Russian equivalent is given in example [3], with the subordinate clause also underlined.

[3] Russian equivalent

Ia rad za tebia, chto ty zdorova i schastliva.  
 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 [2] and [3] reveals that the Tatar utterance using a code-mixed *chto* ‘that’ is almost identical syntactically to Russian – the Tatar is essentially a calque of the Russian with a subordinate clause that comes after the main clause and is introduced by a subordinating conjunction.

In example [4], I give two examples of fully Tatar equivalents. These utterances were both produced by the speaker in example [2] who originally code-mixed *chto* ‘that’. They were produced by him months after [2], when I asked for translation help, claiming that I did not know how say sentence [3] in Tatar. The sentences shown in [4] were the first two offered by the speaker, who upon a few moment’s reflection 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.

[4] Fully Tatar equivalents

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

Sin taza-sau häm kör küngelle bulgan öchen min(bik) shat.  
 you healthy and cheerful mood-with be-PAST.PART for I (very) glad

In both of the sentences in examples [4], the (underlined) subordinate clause is an embedded pre-head clause, as opposed to the Russian-style subordinate clause, which comes after the main clause, as found in [2]. In addition, the clause is not subordinated by a subordinating conjunction, as it is in the code-mixed case and in Russian, but instead with a postpositional phrase or a past participle in the dative. These are significant changes in morphosyntactic realization.

### 6.1.2 Coordinating discourse markers

Two of the Russian coordinating discourse markers code-mixed into Tatar are:

- *i* ‘and’ (conjunction)
- *tozhe* ‘as well’ (particle)

All code-mixed Russian 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. The code-mixing of these coordinators can cause morphosyntactic changes – for example, Russian *tozhe* is used once to coordinate, while its Tatar equivalent is an enclitic used at least twice, once with each element being coordinated.

### 6.1.3 Contrastive discourse markers

Two code-mixed contrastive discourse markers are:

- *ili* ‘or’ (conjunction)
- *no* ‘but, however’ (conjunction)

Example [5] shows a pragmatically typical use of *no* ‘but, however,’ highlighting the negation of a proposition that might otherwise be expected.

[5] Code-mixed *no*

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

**Kstati** anda gel bu sayt-nï – Yuldash-nï – reklamal-ïy-lar.  
**By the way** there always this site-ACC – Yuldash-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.’

Note that this utterance contains a second code-mixed discourse-pragmatic word, *kstati* ‘by the way’. I have kept this discourse-pragmatic word, which is used for metacommentary, to demonstrate just how frequent the production of these code-mixed words can be: it is common to find one code-mixed word per sentence for each sentence in an utterance produced in Tatar-preferred style.

### 6.1.4 Metacommentary

Three of the Russian words code-mixed to produce metacommentary on text are:

- *koroche* ‘in brief’ (adverb)
- *kstati* ‘by the way’ (adverb)
- *slushai(te)* ‘listen’ (verb in imperative)

The words code-mixed for metacommentary have the same discourse uses in Tatar as in Russian. 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. In [6] we see an example of code-mixed *koroche* ‘briefly’, and in [7] code-mixed *slushai* ‘listen’.

[6] 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.**'

*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).

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 [7] 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.

[7] Code-mixed *slushai*

A:    **Slushai,** sin gitara-ng-mī                      küter-ep              kil-ä  
listen      you guitar-2PS.POSS-?PART      carry-2<sup>nd</sup>GER      come-1<sup>st</sup>GER

al-miy-sing-mī?  
take-NEG.PRES-2PS-?PART

F:    Gitara...  
guitar

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:    Guitar...

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

This example includes use of the Russian epistemic *mozhet* 'might', which I will discuss in section 6.3.4. The code-mixing of *slushai* is not syntax-altering, while code-mixing of *mozhet* is.

## 6.2 Interactional performatives

The second major grouping of code-mixed discourse pragmatic words are interactional performatives. There are four such words code-mixed in Tatar-preferred style:

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

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

## 6.3 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.

### 6.3.1 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 include:

- *tochno* ‘precisely, exactly’
- *molodets* ‘well done!’
- *voobshche klassno* ‘totally cool’

Example [8] shows a typical use of *tochno* ‘exactly,’ code-mixed to express positive evaluation.

[8] 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.

### 6.3.2 Negative evaluation

Code-mixed words of negative evaluation are also interjections, e.g.:

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

### 6.3.3 Hedges

Three of the Russian adverbs used as hedges in Tatar-preferred style are:

- *okolo* ‘around’ (adverb)
- *pochti* ‘almost’ (adverb)
- *prosto* ‘just’ (adverb)

The code-mixing of *okolo* and *pochti* results in a word-order change: these adverbs precede the word they are modifying, both in Russian and when code-mixed, while their Tatar equivalents follow the modified word.

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 [9]:

[9] Code-mixed *prosto*

Minem	fiker-em-che,	bezgë	(latin-nï	yaklau-chï-lar-ga)	<b>prosto</b>
My	thought-1PS.POSS-ADV,	we-DAT	(latin-ACC	defending-AGT-PL-dat)	just

sabir	bul-ïrga	kirëk,	äkren	genä	latin-nï	gamäl-gä	kert-ergä,
patient	be-INF	necessary,	slow	only	latin-ACC	usage-DAT	enter-INF,

a	kart	babay-lar	10-15	yïl-da	ülep	bet-ergä	tiesh-lär.
and	old	grandpa-PL	10-15	year-LOC	die-2 <sup>nd</sup> ER	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.’

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

### 6.3.4 Epistemics

Young Tatars code-mix two Russian words as epistemics, that is, to evaluate probability:

- *mozhet* ‘might, maybe, possibly’ (verb, 3ps only)
- *naverno* ‘probably, most likely’ (adverb)

The verb *mozhet* appears in the third-person singular form only. An example of code-mixed *mozhet* was seen in [7] above, and another can be seen in example [10]:

[10] Code-mixed *mozhet*

Ike-nche-dän, min **mozhet** yagimlï 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?’

The Tatar equivalents of *mozhet*, *bälki* 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 [7]). The verb *moch* ‘to be able’ appears to be code-mixed only in this epistemic usage, and in the fixed third-person singular form: it is not used in its root meaning as a conjugated auxiliary verb expressing ability.

### 6.4.5 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 tend to precede the words or propositions being modified, and thus are congruent in their word order.

Two evaluations of time that are code-mixed into Tatar are:

- *uzhe* ‘already’ (adverb)
- *skoro* ‘soon’ (adverb)

Example [11] shows a typical use of *skoro* ‘soon.’

[11] Code-mixed *skoro*

Böten khalik-tan ser it-ep sakl-ïy-lar. Läkin minga  
entire people-ABL secret do-2<sup>nd</sup>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.’

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

Two code-mixed evaluations of degree are:

- *dazhe* ‘even’ (particle)
- *slishkom* ‘too, overly’ (adverb)

Some of these evaluators of degree are words that evaluate the level of the characteristic quality being expressed, while others are used to evaluate the relationship to reality or the general applicability of the proposition in question. In [12] 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.

[12] Code-mixed *obychno*

Kız-lar	tämle	äyber-lär	peshher-ep	kil-ä-lär.	Ä	min	<b>obychno</b>
Girl-PL	tasty	thing-PL	cook-2 <sup>nd</sup> GER	come PRES-3PP.	And	I	<b>usually</b>
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]...’

## 7. Non-content words and morphemes *not* code-mixed

If we exclude from consideration the four code-mixed nouns and verbs found in my corpus of utterances produced by these highly competent Tatar speakers in Tatar-preferred style, then all of the remaining code-mixed words fit into one of the three discourse-pragmatic categories listed above. This is one 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. Interrogatives too are not code-mixed, with the one exception of *kak* ‘how,’ which is used only in its evaluative discourse-pragmatic function, and not 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 not 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) as “prototypical” non-content morphemes) and inflectional and derivational morphology.

## 8. Some comparable findings

In this section, I will give a brief sampling of comparable code-mixing found in similar sociolinguistic situations; a more in-depth examination of such findings, along with a diachronic model of contact-induced structural innovations that takes into account the role of discourse-pragmatic code-mixing, can be found in Wertheim (2003b).

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). These English discourse markers are found in otherwise completely Irish discourse, and Table 1 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 1. English discourse markers found in Irish radio performance*

English discourse marker	Approx. 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
I mean	<i>to est'</i>	yes
now	<i>seychas, teper'</i>	no
y'know	none	n/a

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, eight of the nine code-mixed English discourse

words have Russian equivalents, and seven out of eight of these Russian equivalents are found code-mixed in Tatar.

Further parallels in discourse-pragmatic code-mixing can be found closer to home (home for Tatars, that is) in nearby Uzbekistan. As in Tatarstan, the post-Soviet Uzbek government has engaged in the promotion of the Uzbek language, and a discourse of purity is a prominent language ideology (Baran, 2000: 20). Baran documented twelve Russian words and phrases not found in written Uzbek but used in an “informal register” of “colloquial Uzbek,” even among Uzbek-dominant speakers who almost never code-switched (ibid., 22). These words are found in Table 2, where it can be seen that 10 of the 12 words found in her corpus are code-mixed in Tatar.

*Table 2. 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>prosto</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)

These words are analyzed by Baran as function word borrowings, but her data suggest that these Russian discourse-pragmatic 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.

A final relevant example is that of Russian and Polish discourse-pragmatic words in Karaim, an endangered Turkish language found in Lithuania and Crimea (Csató, 1998, 1999a, 1999b, 2000a, 2000b, in press). The Slavic discourse-pragmatic words found in present-day 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):

Table 3. Russian and Polish discourse-pragmatic words found in Karaim

Russian/Polish discourse-pragmatic word	Found in Tatar?
<i>i</i> ‘and’	yes
<i>a</i> ‘and, but’	yes
<i>nu</i> ‘so, well’	yes
<i>vot</i> ‘there’	yes
<i>uzhe</i> ‘already’	yes
<i>móc</i> ‘to be able’	yes
<i>okolo</i> ‘around’	yes
<i>to</i> ‘then’	no(t yet)
<i>nu to</i> ‘well then’	no(t yet)

In examples [13] and [14], we find see sentences in which the only foreign elements are Russian discourse-pragmatic words; note that their usage and placement is identical to Tatar code-mixing of these very words.

[13] 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 70<sup>th</sup> year.” (Csató 2000a: 681)

[14] Karaim borrowing of Russian *okolo*

**Okolo** b’ir afta-nin Trox-ta kal-ï-m.  
**about** one week-GEN Trakai-LOC stay-AOR-1PS  
 ‘I will stay **about** a week in Trakai.’ (Csató 1999b: 55)

Seven of the nine Slavic discourse-pragmatic borrowings found in Csató’s work are code-mixed in Tatar-preferred style.

## 9. Conclusions

Individual Tatar linguistic performance is set against a purist backdrop with explicit ideals of “pure” and “literary” Tatar. “Pure” Tatar can be equated with “de-Russified” Tatar, and is a performance style used as a prominent part of the construction of Tatar sociolinguistic identity. When Tatars who use this de-Russified style are in a more relaxed in-group setting where a performance style would be inappropriate but Tatar has still been established in some way as the preferred code, they speak in “Tatar-preferred” style. This style is characterized by the seemingly unconscious code-mixing of Russian discourse-pragmatic words. These words fall into three major discourse-pragmatic categories, and are used to mark discourse structure, as interactional performatives, and to express speaker evaluation and stance. In other words, young bilingual Tatars appear to be, to a certain extent, organizing their Tatar discourse in Russian. This is not an isolated phenomenon, but one that has been documented by fieldworkers in similar

sociolinguistic situations, where bilingual speakers of minority languages in a variety of contact situations have been found to code-mix discourse-pragmatic words, even when they are not code-mixing other dominant-language words. This code-mixing appears to be the point of entry for discourse-pragmatic words that become conventional borrowings, as can be seen in the parallels between Slavic code-mixing in Tatar and Slavic discourse-pragmatic borrowings found in Karaim (further parallels in other contact situations can be found in Wertheim 2003b). Code-mixed or borrowed words retain their grammatical requirements, and this leads to structural changes in the minority language; the code-mixing of Russian words leads to a “composite” Tatar that has Russian lexical, morphological, and syntactic elements.

This discourse-pragmatic code-mixing and its effects on the grammatical structure of certain styles of Tatar performance demonstrates that both social and pragmatic factors must be taken into account in order to more fully understand structural linguistic change. Tatar linguistic performance and the Russian-influenced structural innovations found in this performance can only be understood in the context of both interpersonal interactions and higher-level social structures, ideologies, and practices – in particular, Tatar nationalism and the discourse of purity that is associated with it.

## References

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., V. A. Plungian, and E. V. Rakhilina. (1993). *Putovoditel' po diskursivnym slovam russkogo iazyka*. Moskva: “Pomovskii i partneri”.

Blakemore, Diane. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.

Cameron, Deborah. (1995). *Verbal Hygiene*. London/New York: Routledge.

Cotter, Colleen M. (1996). *Irish on the air: Media, discourse, and minority-language development*. University of California, Berkeley dissertation.

Csató, Éva Ágnes. (1998). Über die finiten Verbformen des gesprochenen Nordwest-Karaimischen. In *Turkologie Heute. Tradition und Perspektive. Materialien der Dritten Deutschen Turkologenkonferenz*, Leipzig, 4.-7. Oktober 1994 (pp 47-53). Wiesbaden: Harrassowitz.

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 25<sup>th</sup> Annual Berkeley Linguistics Society meeting* (pp. 54-62). Berkeley: BLS.

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* (pp. 671-699). Berlin: Mouton de Gruyter.

Csató, Éva Ágnes. (2000b). Syntactic code-copying in Karaim. In Östen Dahl & Maria Koptjevskaja-Tamm (Eds.). *The Circum-Baltic languages: Their typology and contacts* (pp. 265-277). Amsterdam: John Benjamins.

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.

Espinal, M. Theresa. (1991). The representation of disjunct constituents. *Language*, 67, 726-762.

Fraser, Bruce. (1996). Pragmatic markers. *Pragmatics*, 6(2), 167-190.

Gil, David. (2001). Escaping eurocentrism: Fieldwork as a process of unlearning. In Paul Newman and Martha Ratliff (Eds.). *Linguistic fieldwork* (pp. 102-132). Cambridge/New York: Cambridge University Press.

Grenoble, Lenore. (1998). *Deixis and information packaging in Russian discourse*. Amsterdam/Philadelphia: John Benjamins.

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. Reprinted in H. Paul Grice. *Studies in the way of words* (pp. 41-58). Cambridge: Harvard University Press.

Maschler, Yael. (1994). Metalanguaging in Bilingual Conversation. *Language in Society*, 23, 325-366.

Myers-Scotton, Carol. (1993). *Duelling languages: Grammatical structure in code-switching*. Oxford: Clarendon 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* (pp. 59-83). Clevedon/Philadelphia: Multilingual Matters Ltd.

Rouchota, Villy. (1998). Procedural meaning and parenthetical discourse markers. In Andreas Jucker and Yael Ziv (Eds.). *Discourse markers: Descriptions and theory* (pp. 97-126). Amsterdam/ Philadelphia: John Benjamins.

Schiffrin, Deborah. (1987). *Discourse markers*. Cambridge/New York: Cambridge University Press.

Searle, John. (1969). *Speech Acts*. Cambridge: Cambridge University Press.

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* (pp. 61-82). Amsterdam/Philadelphia: John Benjamins.

Sperber, Dan and Deirdre Wilson. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.

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.

Wertheim, Suzanne. (2003a). Rethinking the observer’s paradox and data “purity.” In J. Larson and M. Paster, (Eds). *Proceedings of the 28<sup>th</sup> Meeting of the Berkeley Linguistics Society* (pp. 511-521). Berkeley: BLS.

Wertheim, Suzanne. (2003b). *Linguistic purism, language shift, and contact-induced change in Tatar*. University of California, Berkeley dissertation.