

# 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 3<sup>rd</sup> 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: *Makhsatibiñ: Tatarstanning Bässezlege* ‘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.<sup>1</sup>

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.<sup>2</sup> 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

---

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

<sup>2</sup> 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.<sup>3</sup> 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 Schifffrin’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*<sup>4</sup> – 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

---

<sup>3</sup> 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.

<sup>4</sup> Representative uses of these terms are as follows. Schifffrin (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*.<sup>5</sup> 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.

---

<sup>5</sup> 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*),<sup>6</sup> 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

---

<sup>6</sup> 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.<sup>7</sup> 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,

---

<sup>7</sup> 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.<sup>8</sup> 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*

Tatar on-stage style (no Russian)	Tatar-preferred style (unconscious use of Russian discourse-pragmatic words)	code-switching: majority Tatar	code-switching: majority Russian	Russian with Tatar code-mixing	Russian (no Tatar)
--------------------------------------	---	-----------------------------------	-------------------------------------	--------------------------------	--------------------

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

---

<sup>8</sup> 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.<sup>9,10</sup> 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 *molodiets* ‘well done’. Moreover, the matrix language does not need to be consistent for an entire stretch of discourse, and can

---

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

<sup>10</sup> 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.<sup>11</sup> (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

---

<sup>11</sup> 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?”<sup>12</sup> 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

---

<sup>12</sup> In Tatar, *Sin kbäzger duslarıng belän yaki tanışblarıng belän tatarcha söyläsbäsäng. 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 *lektsiia* **chita-iu**  
Ok, I now *lecture* **read-PRES.1ps**  
'Ok, now I'm **giving** you a *lecture*.'<sup>13</sup>

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 *lektsiia* ‘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

---

<sup>13</sup> 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 tanışbular 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.<sup>14</sup> 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	küstir-miycha
pure	Tatar –	it	Russian	word-PL-POSS-ACC	insert- NEG.PRES.GER
söyläshü,	tınglau-chi-lar-ga	<b>priyatno</b>	bul-irga	tiesh	tıngla-rga.
conversation	listening-AGT-PL-DAT	<b>pleasant</b>	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 20<sup>th</sup> and 21<sup>st</sup> 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 10<sup>th</sup> century Bolgar conversion to Islam. As superstrates,

---

<sup>14</sup> 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-zï 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-nï barla-u, boringï  
 this self-1PP-GEN historical source-PL-ACC verify-V.NOUN ancient

baba-lar-ïbüz jir-en-ä säjdä kil-ü, alar-nïng  
 grandfather-PL-1PP earth-POSS-DAT bow do-V.NOUN they-GEN

rukħ-ïn-a bagışhla-p doga uk-u da **bit**.  
 soul-POSS-DAT devote-2<sup>nd</sup> GER prayer read-V.NOUN also **EMPH.PART**

**Ä** shushï 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(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 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.’ –A.

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

[3.7] Russian equivalent

Ia    rad    dlia    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 [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 *dlia* 'for'. Note that morphologically, the Russian adjectives all express gender (masculine for *rad* 'happy', and feminine for *zdorova* 'healthy' and *schastliva* '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 *cbto* 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 *cbto* 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 cbto* ‘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 *chto*. These eleven methods of subordination are listed in Appendix B in the entry for *chto*.

Utterances using a code-mixed *chtoby* ‘in order to, that’ are also realized with a different clausal structure than if produced in Tatar only. As with *chto*, a Tatar sentence using the word *chtoby* as a subordinating conjunction will have a subordinate clause that follows the main clause, Russian-style. Additionally, when *chtoby* 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 *öchen* ‘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 *-üina*), and (4) a verb in the 3<sup>rd</sup> person singular imperative form, *-sîn*.<sup>15</sup>

Subordinate clauses using Russian *potomu chto* ‘because’ to express causality are syntactically identical to those using the subordinating conjunction *chönki* ‘because.’ While constructions using *chönki* are considered to be “pure” Tatar, *chönki* 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 *chönki*, 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 chto*.

---

<sup>15</sup> All of the suffixes and enclitics described in this chapter have underspecified vowels that are subject to vowel harmony. The 3<sup>rd</sup> person imperative suffix *-sîn* 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 kürü* 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.-ni da chakir klub-ka. Anga **tozhe** ber kiz 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 *tozbe* 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	Dulkïn-ga	shaltïrat-ip	kara	äle.	<b>Ili</b>	Kazan-da	ul	tagïn
You	Dulkïn-DAT	call-2 <sup>nd</sup> GER.	see	PART.	<b>Or</b>	Kazan-LOC	he	again
kaychan	bulachag-i-n		belep		kara.			
when	be-FUT.DEF-POSS-ACC		know-2 <sup>nd</sup> GER.		see-IMP.			

‘Try and call Dulkïn [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-mïy-m.  
 This week-LOC amateur-PL concert-POSS **but** I sing-NEG.PRES-1PS.

**Kstati** anda gel bu sayt-nï – Y.-nï – reklamal-ïy-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 ungayında* ‘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)
  
- *vot* ‘there’
- *vot, 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 *vot, 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*

<b>Koroche</b>	shaltirat-kach	söylesh-äbez.	<b>Poka.</b>
<b>Briefly,</b>	call-PAST.PART	speak- PRES.1PP.	<b>Bye.</b>
<b>'In short, after you call, we'll talk. Bye.'</b> –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 doshman* ‘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	küz-lar-ï	tügel.	Mishär	tel-e-n,
P.	we	EMPH.PART	Mishar	girl-PL-POSS	not.	Mishar	language-POSS-ACC

<b>to est'</b>	dialekt-ï-n	angla-miy-biz	bit.
<b>that is</b>	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	<b>naprimer</b>	Idel	tämäke-se-n
he	in reality	bourgeois	I	<b>for example</b>	Idel	tobacco-POSS-ACC
tart-a-m,		ä	ul	Parliament	tart-a.	
smoke-PRES-1PS		and/but	he	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*

<b>Vot</b>	singa	<b>tipychnyi</b>	kön.
<b>There</b>	you-DAT	<b>typical</b>	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 *mi* 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-mi küter-ep kil-ä al-miy-sing-mi?  
**listen** you guitar-2PS.POSS-?PT carry-2<sup>nd</sup>GER come-1<sup>st</sup>GER take-NEG.PRES-2PS-?PT

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.

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 buligiç* (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.<sup>16</sup>

Perhaps the same is true for the use of Russian *pozhaluista* ‘please’ rather than Tatar *zinbar* or *zinbar 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*: *zinbar* 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*. *Içvini(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 itegeç* (plural). For all of the words in this category, “their

---

<sup>16</sup> 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      Kön-e      **kak?**  
Memory    day-POSS   **how?**  
‘**How** was Memory Day?’ –J.

As described in Chapter One, *Khäter Köne* ‘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!’
- *točno* ‘precisely, exactly’
- *voobščee klassno* ‘totally cool’

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

[3.25] Code-mixed *točno*

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

A: **Točno.** ‘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 *voobshe 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!’
- *zhal’ko* ‘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)	<b>prosto</b>	
My	thought-1PS.POSS-ADV,		we-DAT	(latin-ACC	defending-AGT-PL- DAT)	<b>just</b>	
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	yil-da	ülep	bet-ergä	tiyesh-lär.
and	old	grandpa-PL	10-15	year-LOC	die-2 <sup>nd</sup> 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	<b>pochti</b>	nol	kaloriya	genä.			
and	<b>almost</b>	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-2<sup>nd</sup>GER leave-NEG-PAST.DEF-1PP ok EMPH PART

zaman-nar dürt jir jirla-p uya-t-ti-lar.  
time- PL four song sing-2<sup>nd</sup>GER 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 *kitmüdek* 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 2<sup>nd</sup> 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 <sup>st</sup> 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 <sup>nd</sup> 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*, *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 [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 <sup>nd</sup> GER	preserve- PRES-3PP	but	me-DAT
informer-lar	jitker-de-lär.	<b>Skoro</b>	min	dä	Moskva-ga	bar-miy-m.
informant-PL	lead-PAST.DEF-3PP	<b>soon</b>	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-2 <sup>nd</sup>	GER	look-IMP	PART	there	I you-DAT
khat	yaz-gan	i-de-m.	Mömkin	bul-sa	<b>srochno</b>	gïna		
letter	write-PAST.PART	be-PAST-1PS.	Possible	be-COND	<b>urgently</b>	PART		
jawap	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, *achigïp*, 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 *achigïp* 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]

- *dazhe* ‘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 <sup>nd</sup> GER
<b>dazhe</b> khat-ta <b>even</b> 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 vsegda* ‘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 *absolint* 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	<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]...’ –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	jür-chi	eis of beis,	doktor alban,	<b>v obshchem</b>	min
love-PAST.PART	song-AGT	Ace of Base	Doctor Alban	<b>in general</b>	I
jür-ga	kara-p	yarat-a-m.			
song-DAT	look-2 <sup>nd</sup> 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	<b>inogda?</b>
look-IMP	EMPH.PART	you	read-PRES-2PS-?PT	intertat-ACC	<b>sometimes</b>

‘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	<b>na samom dele</b>	sikher bar.
and/but	even	<b>in reality</b>	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 *byl* ‘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.