

Mixed languages: Patterns of lexification of the core vocabulary

Ekaterina M. Gridneva

National Research University Higher School of Economics (HSE University)
(Moscow, Russia); egridneva@hse.ru

Abstract. This paper is an analysis of mixed languages in terms of the sources of lexicalization of their core vocabulary. Data from the following mixed languages are considered: Michif, Medny Aleut, Media Lengua and Gurindji Kriol. Previous studies devoted to the typology of mixed languages from the combined perspective of structure and lexicon are also examined [Bakker 2017: 218], [Meakins 2018: 6], [Muysken 2008: 211]. This paper focuses on the core lexicon alone. First, I consider claims concerning the whole vocabulary of mixed languages. Next, using Swadesh 100 and 207 word lists as the basis of comparison, I examine to what extent the generalizations about the lexicon of the mixed languages also apply to the core lexicon. I present the patterns of the lexification of the core vocabulary and conclude that the observations about lexification of mixed languages also hold for their core lexicons. The comparison leads to some findings with respect to lexical doublets (i.e., two lexical items from different lexifier languages corresponding to one concept in the core vocabulary), with more lexically mixed languages showing a higher number of doublets. I conclude that focusing on the core vocabulary does not resolve the problematic status of mixed languages in terms of the conventional historical comparative analysis. Additionally, this paper presents and tests the synchronic hypothesis attributing the number of doublets to the “more mixed nature” of one mixed language. In other words, there is a direct correlation between the quantity of doublets and the relevant language’s lexis in terms of even distribution between the lexifier languages. The more a language is “mixed”, the more doublets it has.

Keywords: mixed languages, lexifier languages, core vocabulary, Michif, Medny Aleut, Media Lengua, Gurindji Kriol, Swadesh list.

Лексификация смешанных языков: исследование базового словаря

Е. М. Гриднева

Национальный исследовательский университет «Высшая школа экономики»
(Москва, Россия); egridneva@hse.ru

Аннотация. Статья посвящена описанию моделей лексификации смешанных языков с точки зрения их базового словаря. Исследуются следующие языки: мичиф, алеутско-медновский язык, медиа ленгуа, гуринджи криол. Рассматриваются предыдущие исследования в области типологии смешанных языков, а также сопоставляются общие положения об источниках лексики в конкретных смешанных языках в целом с результатами анализа лексификации значений из списка Сводеша. Предлагается классификация смешанных языков в зависимости от их лексического состава и распространенности дублетов (наличие двух лексем из разных языков-лексификаторов, выражающих одно и то же значение).

Ключевые слова: смешанные языки, мичиф, алеутско-медновский язык, медиа ленгуа, гуринджи криол, язык-лексификатор, список Сводеша.

1. Introduction

Thomason [1997: 21] suggests that “all languages are mixed in a weak sense: there are no natural human languages in which foreign material is wholly lacking.” The consequences of a language contact can, however, differ. One can distinguish three types of language contact outcomes: i) language maintenance, ii) language shift, and iii) a new language creation [Matras 2000: 80]. Three types of contact languages are further distinguished: pidgins, creoles and mixed languages [Heine, Kuteva 2005: 20]. Mixed languages as a special type of contact languages were identified as early as in [Thomason and Kaufman 1988], [Bakker, Mous 1994] and [Thomason 1997]. Unlike pidgins and creoles, mixed languages normally emerge from combining grammar, phonology and lexicon of two identifiable source

languages in a situation of community bilingualism [Matras 2000: 80]. Creoles usually — though not always — have one dominant, clearly identifiable lexifier language. For example, Tok Pisin was drawing its vocabulary primarily from English, with a much smaller input from German, Malay, Portuguese and Austronesian languages [Mühlhäusler 1981: 36]. Unlike pidgins but similarly to creoles, mixed languages are also linguistically complex and are used in all communicative domains [Bakker, Matras 2013: 159]. From a social perspective, mixed languages emerge with the appearance of a new community (e.g., Métis in Canada) [Bakker 1997: 12] as a result of immigration, which may lead to mixed marriages (e.g., Ma'a, Michif, Medny Aleut), or cultural incursion [Meakins 2018: 12]. In this regard, the genesis of mixed languages is explained not by a merely communicative need. From the moment when they arise, they serve as markers of the identity of a particular group [Golovko 1994; Bakker 1997; Meakins 2014: 393].

There are various models of mixed language formation. For example, Michif emerged as a result of language intertwining which takes place when the vocabulary from one language and the grammatical system from another are combined together [Bakker 1997: 203]. Though Michif represents a far more complex case of language intertwining, and stands out among all other similar instances due to certain typological properties of Cree, the process of its emergence is evident [Bakker 1997: 213]. Michif “shows the existence of French lexical stems with Cree grammatical affixes. When the Cree affix is not used, there is a French element that functions unlike its French source but exactly like the Cree element it replaces, with only few exceptions” [Bakker 1997: 247]. Relexification is a situation where a vast body of the lexicon (e.g., stems) of one language is replaced by lexical items from another language, as is the case in *Media Lengua* [Muysken 1981]. Finally, there is a code-switching formation model [Meakins 2005] where speakers shift from one language to another (e.g., as in *Gurindji Kriol*).

The first detailed account of a mixed language, Michif, appeared in Bakker's [1997] *A Language of Our Own*. This paper mostly builds on the following descriptions of mixed languages: Michif [Bakker 1997], Medny Aleut [Golovko 1994], *Media Lengua* [Muysken 1997] and *Gurindji Kriol* [Meakins 2005].

My aim is to look at the patterns of lexification of the core vocabulary, to review the general claims regarding lexification of mixed languages, and examine whether these also hold for the latter's core vocabulary. Further, I compare Swadesh 207 to Swadesh 100 lists [Swadesh 1971: 283] as less versus more stable vocabulary to see whether they show similar behavior. By subtracting the words included in the Swadesh 100 list from the Swadesh 207 list, I obtained two lists to be hereinafter denoted as the more stable vocabulary (Swadesh 100 list) and the less stable vocabulary (Swadesh 207 list minus the subtracted Swadesh 100 list). Additionally, I present some findings regarding doublets in mixed languages, i.e. a lexification pattern where two lexical items for a particular lexical concept, while originating from two donor languages, show more or less equal frequency.

This article does not attempt to resolve the problematic status of mixed languages in terms of historical and comparative analysis, its aim being to merely assess the empirical grounds for such an analysis. Nevertheless, a brief introduction into the problems of genealogical classification of contact languages is needed here. In his discussion of the dual affiliation of the creole languages of the Caribbean, Taylor argues that, while inheriting their vocabulary from one 'parent', they borrow their grammar from the other. Putting more emphasis on structure, he proposed to define the relationship of the first type as genetic, and the second, as deep (basic) [Taylor, 1956]. A different view holds that phonetic correspondences between creoles and their lexifiers cannot be mistaken by the traditional comparative method for a genetic relation since the phonetic correspondences between them do not result from regular sound change. Also, glottochronological calculations of the split-off date for mixed languages can be largely off the mark and vary highly depending on which of the lexifier languages is considered. Bakker [2000] insists that "an 86 percent retention rate in 1000 years would suggest a split-off around two millennia ago (from a French viewpoint) or three millennia (from a Cree viewpoint), while we know that the Michif language cannot be that old". Moreover, the majority of the words in the Swadesh lists are nouns, which can lead to some invalid conclusions for languages with a lexifier split along part-of-speech categories. Michif nouns and adjectives primarily originate from French,

while verbs come from Cree. This does not mean that the language is more French- than Cree-based, even if there are more nominal than verbal concepts in the Swadesh list (or, for that matter, in the lexicon at large) [Bakker 2000: 600]. Consequently, the glottochronological approach based on Swadesh lists may deliver results that are invalid in terms of the conventional historical comparative analysis. This makes it important to see whether the claims made about the lexicon of an individual mixed language as a whole also hold for its core vocabulary, often focused by conventional historical analyses, and is thus the aim of the present study.

The Swadesh 100 lists for Michif and Media Lengua, respectively, were made available in Belikov [2006: 88] and Bakker and [2000: 601]. The Swadesh 207 lists for Medny Aleut and Michif were collected by me based on the available dictionaries. For Medny Aleut, I used the dictionary by E. Golovko and N. Vakhtin [1994], for Media Lengua, lists compiled by Pieter Muysken, and for Gurindji Kriol, lists by Felicity Meakins.

In the next section, I provide a context for this study by overviewing older attempts and approaches towards a typology of mixed languages. *Section 2* is followed by sections with brief overviews of the respective mixed languages addressed and findings of their analyses: Medny Aleut (*Section 3*), Michif (*Section 4*), Media Lengua (*Section 5*) and Gurindji Kriol (*Section 6*). *Section 7* presents my observations on doublets in different languages and *Section 8*, a discussion of their implications.

2. Mixed language typologies

This section proposes a survey of earlier typologies of contact languages. All classifications of mixed languages surveyed invariably address their lexical composition. Thus, Bakker [2015] provides a classification of mixed languages by the sources of their linguistic structures and isolates the following types. Type 1 or G-L mixed languages includes idioms where the source of the grammar (G language) differs from the source of the lexicon (L language). Thus in Media Lengua, Spanish lexical roots are embedded in Quechuan morphology. Type 2 are F-R languages where in Sri Lanka

Portuguese, for example, the typological frame F (an abstract grammatical frame) comes mostly from Tamil, while roots R (or lexicon), from Portuguese. The main difference between the first and the second type is that bound morphemes in the first type remain identical to those of the G language, while in the second type, both bound morphemes and roots etymologically derive from the same language. Type 3 includes N-V languages where nouns come from one language and verbs, from the other as in Michif, where nouns come from French and verbs from Cree. Gurindji Kriol is an example of a V-NN language (Type 4), where verbal phrases pattern after Kriol verb phrases and converb and noun phrases, after those of Gurindji, while the lexicon of the noun phrase is a mixture of both Gurindji and Kriol lexis. Medny Aleut (Type 5) is an L-INF language where basically, the noun phrase and verbal roots are Aleut, while the finite verb inflections come from Russian. Finally, in LL languages (Type 6), the basic lexicon is mixed. These are mostly creoles and pidgins (e.g., Berbice Creole).

Muysken [2008] proposes a different classification based on structural frames (grammar — lexicon, verb — noun) and including a sociolinguistic perspective (the original community language versus the introduced language). He divides mixed languages into four types. The ‘classical’ type (e.g., Media Lengua) represents languages where the morphosyntax and functional categories come from the original community language and the lexicon, from the other language. The ‘split’ type (e.g., Michif) includes languages with the verbal system coming from the original community language and the nominal system from the introduced language. The ‘split reverse’ type is represented by Medny Aleut and Gurindji Kriol. In these languages, the verbal system does not arrive from the original community language, while the nominal one does. For example, in Medny Aleut the finite verb morphology is Russian (the introduced language) while the nominal inflection is Aleut (the original community language). Finally, the ‘reverse’ type (e.g., Angloromani) shows a grammar-lexicon divide, with the lexicon arriving from the original community language.

Meakins [2018] classifies mixed languages into 1) *(L)exicon-(G)rammar languages* where languages fall into: those providing the grammatical structure and those contributing to the lexicon (e.g., Media Lengua, Angloromani); 2) *structural mixes* where both languages contribute to both

Table 1. Typologies of mixed languages

	Bakker [2015]	Muysken [2008]	Meakins [2018]
Michif	N-V (nouns: French; verbs: Cree)	Split type (verbal system: original community language, i.e. — Cree; nominal system: introduced language, i.e.: French)	Structural mixes: both of the source languages contribute to the structure and lexicon of the resultant mix language.
Medny Aleut	L-INF (noun phrases and verbal roots: Aleut; finite verb inflection: Russian)	Split reverse type: Medny Aleut and Gurindji Kriol (verbal system / finite verb morphology: introduced language; nominal system / nominal inflection: original community language)	
Gurindji Kriol	V-NN (verb phrase: Kriol; converbs and noun phrases: Gurindji BUT lexicon in the noun phrases can be both Gurindji and Kriol)		
Media Lengua	G-L (grammar: Quechua; lexicon: Spanish)	Classical type (morphosyntax and functional categories: original community language, i.e. Quechua; lexicon: introduced language, i.e. Spanish)	L(exicon) — G(rammar): morpho-syntactic frame is of the ancestral language or Quechua, while 90 % of its stems are by Spanish forms (introduced language)

the lexicon and grammar of the new language (e.g., Medny Aleut, Michif and Gurindji Kriol); and 3) *converted languages* (e.g., Sri Lanka Malay) where the ancestral language maintains its lexicon but undergoes a complete restructuring of its morphosyntax based on the introduced language.

To sum up, all the typologies are based on structural-lexical features and historical perspective. In the current study I am going to check whether the claims about specific mixed languages these typologies make hold for their basic vocabulary.

3. Medny Aleut

Medny Aleut is spoken by the Aleuts of Medny (one of the Commander Islands in the south-eastern part of the Bering Sea). Aleuts are the indigenous population of the Aleut Islands. Most Aleuts live in the United States (Alaska), and some in Russia (Kamchatka Krai). Medny Aleut was first encountered by G. Menovshchikov during his 1963 expedition. The language probably emerged in the late XIX century at the time of Medny Island settlement [Golovko 1997] as a result of intensive contacts between several Aleut dialects and Russian. It is difficult to estimate the present number of Medny Aleut. The 1980s report data put it at no more than 10–12 speakers. The language seems to be mostly based on the Attuan dialect originating from Attu, the westernmost Aleut island, and was spoken until the mid-XX century.

Medny Aleut integrates two phonological systems, those of Aleut and Russian, albeit with a greater proportion of Aleut features. Medny Aleut is an agglutinative language just like Aleut from which it also inherits its derivational and nominal inflection systems. Its finite verb morphology is Russian, while the verb roots themselves are largely drawn from the original community language (Aleut), with only a small proportion of the verb roots coming from Russian [Golovko 1994: 115].

Although the vocabulary is of primarily Aleut origin, it includes a large number of Russian words, as stated by Sekerina [1994: 29] who relied on a vocabulary drawn from examples and sample texts from [Vakhtin

1985], [Golovko 1989], and [Golovko, Vakhtin 1990]. The majority of nouns and verbs (61.5 % and 94 %, respectively) come from Aleut. Adjectives do not constitute a separate syntactic category [Sekerina 1994: 24]. A qualitative attribute is expressed by nouns in the possessed form, while a qualitative predicate is expressed by a verb. Personal pronouns and demonstratives come from either Russian or Aleut. Most interrogative pronouns are Aleut. In total, according to Sekerina [1994], 33.5 % pronouns come from Aleut and 66.5 % from Russian. Numerals from 1 to 10 come from Russian and Aleut, with the other numerals borrowed from Russian. Adverbs, negators, particles and conjunctions come from Russian, while the other function words are Aleut (31.5 %) or Russian (68.5 %).

As indicated earlier, this study relies on the less and more stable vocabulary lists (Swadesh 207 and Swadesh 100). Medny Aleut lists were collected

Table 2. Medny Aleut: more and less stable vocabulary lists in natural numbers

	More stable vocabulary				Less stable vocabulary			
	Aleut	Russian	Both	Total	Aleut	Russian	Both	Total
nouns	43	3	1	47	21	4	1	26
verbs	18	0	0	18	31	0	0	31
adjectives	10	0	1	11	14	1	0	15
numerals	0	0	2	2	0	0	3	3
pronouns	0	5	0	5	2	3	0	5
quantifier (all, many, some, few, other)	1	1	0	2	1	1	0	2
interrogative (who, what, where, when, how)	1	0	1	2	1	2	0	3
function words	0	1	0	1	1	3	2	6
Lexifier	73	10	5	88	71	14	6	91

by the author from the available dictionaries. *Table 2* describes the lexification of the core vocabulary based on the more stable vocabulary (Swadesh 100) and less stable vocabulary (Swadesh 207 minus Swadesh 100).

The table above shows that on both counts the bulk of the vocabulary comes from Aleut (83 % and 78 %, respectively). Interestingly, the lexeme ‘father’ has two variants with an Aleut and a Russian root: *ayaxx* and *taatkaxx*¹, respectively, while ‘mother’ is of the Russian origin: *maamkaxx*. Notably, while the Medny Aleut vocabulary, including nouns, mainly comes from Aleut, such basic concepts as ‘mother’ and ‘father’ are partly or fully Russian.

In consistency with the existing typologies, Medny Aleut demonstrates a clear leading lexifier language. While Meakins [2018] classifies Medny Aleut as a structural mix (a split between the lexicon and grammar where “by implication, both languages also contribute to the lexicon of the resultant mix”) [Meakins 2018: 7], the Swadesh lists show the Aleut influence to be much stronger, with 83 % and 78 % of core vocabulary respectively contributed by Aleut.

If we compare these counts to the data in [Sekerina 1994], the results will be only slightly different for verbs (94 % in Sekerina vs. 100 % in both Swadesh lists). The difference is more evident for nouns where Aleut shows 91 % and 81 % in our data and “only” 61.5 % in [Sekerina 1994]. The counts are very similar for pronouns and function words either. Thus, from the core vocabulary perspective, the Medny Aleut lexicon comes from the Aleut language even more consistently compared to Sekerina’s analysis of running texts.

4. Michif

Michif is spoken in some provinces of Canada (Manitoba, Saskatchewan) and in two American states (North Dakota, Montana). The

¹ The Russian word тата /tata/ or тятя /t’at’a/ means ‘father’ in some Russian dialects.

Métis people are descendants of European fur trader fathers (often French Canadians) and Cree-speaking Amerindian mothers. By 2010–2011 data, there were 730 native Michif speakers; also see [Bakker 1997: 5].

There exist a number of approaches to modeling the social circumstances and linguistic mechanisms of the emergence of Michif. The social models address the creation of a new ethnicity, the trade language emergence hypothesis, the organization of the household, etc. The linguistic models include: a) a verb-noun mixing theory whereby the first Métis generation started mixing the two languages by recruiting verbs from one language and nouns, from the other; b) Michif could have resulted from French-Cree code mixing; c) the relexification hypothesis proposes a unique development whereby not the whole vocabulary, but only Cree nouns were replaced by French nouns; and d) a ‘difficult parts’ hypothesis explains that Michif combines the most difficult parts of both languages, i.e. complex French nouns and complex Cree verbs [Bakker 1997: 13].

Michif combines two phonological systems, that of Cree and that of French, with two different phoneme inventories. The noun phrase in Michif is essentially French. Nouns are always accompanied by a French determiner or a possessive pronoun. Most often, nouns have the same gender as in French. Cree demonstratives, animate or inanimate, can be added to the French noun phrase. Adjectives are always French. Though a few adjectives are Cree, they do not represent a morphosyntactic class in Cree, and noun modifiers are expressed either by verbal constructions (relative clauses), or by prefixation to the noun Bakker [1997: 106]. The Michif verb phrase is basically that of Plains Cree. Numerals are always French, except occasional use of the Cree numeral *payyek* ‘one’. Question words virtually all come from Cree. Function words (question words, discourse particles, etc.) are mostly Cree, except for the French articles and prepositions. Most of the categories (property words, pre- and postpositions, adverbs, negation, and conjunctions) are drawn from both languages, with some regional variation. According to the brief sketch of Michif’s lexicon in Bakker [1997: 117]:

- Nouns: 83–94 % French; rest Cree or Ojibwe, English;
- Verbs: 88–99 % Cree; few French verbs; some mixed Cree and French;
- Question words: almost all Cree;
- Personal pronouns: almost all Cree;
- Adverbial particles: 70 % Cree; 30 % French;
- Postpositions: almost all Cree;
- Coordinate conjunctions: 55 % Cree; 40 % French; 5 % English;
- Prepositions: 70–100 % French; rest Cree;
- Numerals: almost all French;
- Demonstratives: almost all Cree;
- Negation: roughly 70 % French; 30 % Cree.

Table 3 shows the counts for the core lexicon based on the Swadesh lists.

Table 3. Michif: more and less stable vocabulary lists in natural numbers

	More stable vocabulary				Less stable vocabulary			
	Cree	French	Both	Total	Cree	French	Both	Total
nouns	0	43	9	52	0	27	1	28
verbs	18	0	1	19	35	0	2	37
adjectives	0	8	6	14	9	7	3	19
numerals	1	1	0	2	0	3	0	3
pronouns	5	0	0	5	4	0	0	4
quantifier (all, many, some, few, other)	2	0	0	2	3	0	0	3
interrogative (who, what, where, when, how)	2	0	0	2	3	0	0	3
function words	1	0	0	1	2	2	0	4
Lexifier	29	52	16	97	56	39	6	101

A comparison of the more and less stable vocabulary lists shows two different pictures. The more stable vocabulary list (the Swadesh 100 list) is predominantly French with 52 French vs. 29 Cree lexical items, while the less stable vocabulary list is led by Cree (56 Cree vs. 39 French items). The explanation may lie in the specific nature of Michif. As mentioned before, Michif shows a split between particular parts of speech whereby, for example, nouns have French and verbs, Cree origins, etc., while the more stable and the less stable vocabulary lists demonstrate a different part-of-speech (nouns vs. verbs) distribution.

The word class distribution corresponds to the general claims regarding the lexicon. Exceptions include adjectives. Comparing the two lists one can see that in the first list adjectives come from French and/or from doublets, while in the second one the majority of adjectives are of almost equally Cree or French origin with only few doublets. In the more stable vocabulary list, the French borrowings include the following lexical concepts: ‘small’, ‘new’, ‘round’, ‘red’, ‘green’, ‘yellow’, ‘white’, ‘black’. The doublets express the following meanings: ‘big’, ‘long’, ‘cold’, ‘full’, ‘good’, and ‘dry’. The less stable vocabulary list includes such French-originated concepts as ‘wide’, ‘thin’, ‘old’, ‘rotten’, ‘straight’, ‘right’, and ‘left’, while the Cree borrowings include ‘thick’, ‘heavy’, ‘warm’, ‘dirty’, ‘sharp’, ‘dull’, ‘wet’, ‘correct’, and ‘far’. While it is difficult to estimate any semantic dependencies in terms of adjectives and their lexifier languages, the terms for colors mostly came from French.

To sum up, the core Michif lexicon does show the lexical patterns suggested by the generalizations about its lexicon but for the less stable adjectives. Additionally, Michif’s core vocabulary patterns by both Swadesh lists fit into the existing mixed language typologies proposed by Bakker [2015], Muysken [2008], and Meakins [2018]. At the same time, as the discussion above shows, Michif presents a stronger structural mixture in its core vocabulary compared to Medny Aleut, while both languages belong to the same type of structural mixes according to [Meakins 2018].

5. Media Lengua

Media Lengua is a mixed language spoken by a Quechuan group known as the Obreros² in the Andean region of Central Ecuador [Muysken 1997: 374]. It is usually described as a contact language with a Quechuan morphosyntactic structure but with almost all content words³ (around 90 %) replaced by Spanish-derived forms. Muysken suggests that it emerged by a process of relexification where Spanish stems replaced their Quechuan counterparts. This process was set off in 1967 by a group of young construction workers in a provincial town who created this language because they, as acculturated Indians, failed to fully identify themselves with either the traditional Quechua culture, or the urban Spanish culture [Muysken 1997: 376]. According to Dikker [2008: 121], Media Lengua “was created by men who had Quechua as their native language but left to work in Spanish speaking areas. When they returned to the communities, they had been using Quechua on an infrequent basis, while having acquired relatively fluent urban Spanish”. She argues that Media Lengua connects the older monolingual Quechuan speaking generation and the younger monolingual Spanish speaking generation.

Muysken [1997] provides a 207 Swadesh list of the core vocabulary for the local Quechua and for Media Lengua (with items, for which the Media Lengua form is unknown, omitted). *Table 4* summarizes his data as counts.

The counts correspond to the general description of the language. While the Swadesh lists confirm that the majority of content words were replaced by Spanish derivatives, there are some exceptions, including:

‘child’	<i>wawa</i>	<i>wawa</i>
‘wife’	<i>warmi</i>	<i>warmi</i>
English	Local Quechua	Media Lengua

² Workers residing in villages near the town of San Miguel de Salcedo.

³ That is, lexical category words like nouns, verbs, adjectives and most adverbs.

Table 4. Media Lengua: more and less stable vocabulary lists in natural numbers

	More stable vocabulary				Less stable vocabulary			
	Local Quechua	Spanish	Both	Total	Local Quechua	Spanish	Both	Total
nouns	2	10	1	13	2	14	0	16
verbs	2	15	0	17	1	12	0	13
adjectives	1	6	1	8	1	5	1	7
numerals	0	2	0	2	0	3	0	3
pronouns	0	4	0	4	0	4	0	4
quantifier (all, many, some, few, other)	0	1	0	1	0	3	0	3
interrogative (who, what, where, when, how)	0	1	0	1	0	3	0	3
function words	0	1	0	1	5	0	0	5
Lexifier	5	40	2	47	9	44	1	54

Notably, almost all Media Lengua prepositions and conjunctions also come from Quechua.

To conclude, the core vocabulary patterns fully correspond to the descriptions of Media Lengua provided in Bakker [2015], Muysken [2008], and Meakins [2018] typologies, with all these authors suggesting that its vocabulary comes from the introduced language (Spanish) while the grammar comes from the original community language (Quechua).

6. Gurindji Kriol

Gurindji Kriol is spoken by Gurindji people in the Victoria River area in northern Australia. Its two lexifiers are Gurindji and Kriol where Gurindji is a Pama-Nyungan language and Kriol is an English-lexified

creole. Gurindji Kriol is a linguistic outcome of the contact between non-Indigenous settlers and Gurindji people, the traditional owners of the land. When, looking for good pastures, the colonists seized the traditional Gurindji lands in the mid-XIX century, Gurindji Kriol emerged in the course of contacts between the white pastoralists (English speakers) and the conquered Gurindji. Following the initial period of conflict in the late XIX and early XX centuries, many of the Gurindji found themselves working on cattle stations as kitchen hands or stockmen. This gave rise to a lingua franca, an English-based pidgin which later developed into Kriol. McConvell [1988] found that in the 1970s code-switching between Kriol and Gurindji was the dominant language practice among the Gurindji people. "... [T]his code-switching and a certain amount of leveling between Gurindji and closely-related neighboring languages such as Ngarinyman and Bilinarra provided fertile ground for the formation of the mixed language" [Meakins 2013: 131].

Most of the NP structure including case and derivational morphology originated from Gurindji. Kriol supplies much of the verbal grammar including tense and mood auxiliaries, as well as transitive, aspect and derivational morphemes. According to Meakins [2013], the phonological systems of Gurindji and Kriol are quite similar because the cattle station pidgin developed in the Victoria River District under the influence of Gurindji. Gurindji Kriol has two separate phonological inventories, one

Table 5. Distribution of Kriol and Gurindji elements in Gurindji Kriol [McConvell, Meakins, 2005: 11]

Solely Kriol		←————→		Solely Gurindji	
non-emphatic pronouns	temporals	adjectives	possessive pros	emphatic pronouns	
demonstratives	directions	N-people	in-law kin	emphatic demonstratives	
V-basic	colours	N-animals	N-body parts	V-bodily	
conjunctions	close kin	N-food	V-state	functions	
counting numbers		fire, cook	V-motion	grandparent kin	
		quantifiers	V-impact	N-plants	
		interjections	adverbs		

coming from Gurindji and the other from Kriol; same applies to the syllable structure and phonological processes. Stress is the word initial for both Gurindji- and Kriol-borrowed items.

Table 6. Gurindji Kriol: more and less stable vocabulary lists in natural numbers

	More stable vocabulary				Less stable vocabulary			
	Gurindji	Kriol	Both	Total	Gurindji	Kriol	Both	Total
nouns	27	7	15	49	5	16	8	29
verbs	6	3	10	19	9	23	3	35
adjectives	2	8	3	13	4	15	2	21
numerals	0	0	2	2	0	2	1	3
pronouns	0	0	5	5	0	0	5	5
quantifier (all, many, some, few, other)	0	1	1	2	0	2	1	3
interrogative (who, what, where, when, how)	0	1	1	2	0	2	1	3
function words	0	1	0	1	0	3	3	6
Lexifier	35	21	37	93	18	63	24	105

Table 6 shows the lexifier language and word class distribution in Gurindji Kriol. The more stable vocabulary includes doublets (37 lexical items), with the other lexical items coming, respectively, from Gurindji (35 items) and Kriol (21 items). In contrast, the less stable vocabulary is mostly from Kriol (63 lexical items), the other contributions representing doublets (24 lexical items), and Gurindji (18 lexical items). An analysis by the word class distribution reveals different results as well. In terms of the more stable vocabulary, the lexifier language is Gurindji for nouns, and Kriol for adjectives, while verbs represent doublets. In consistence with the less stable vocabulary list, the majority of nouns, verbs, and adjectives come from Kriol. Interestingly, Gurindji Kriol has the highest number of doublets compared with the other mixed languages addressed here.

Meakins [2018] defines Gurindji Kriol as a structural mix where both source languages contribute to its structure and lexicon. The Swadesh lists corroborate this conclusion for Gurindji Kriol’s core vocabulary.

7. Doublets

In this section, I consider doublets in mixed languages. Meakins [2013: 137] defines doublets as “synonymous forms from both languages which are used interchangeably, depending on a number of sociolinguistic factors including group identification and the age of the addressee”. For example, the Gurindji form *tipart* ‘jump’ may be chosen if the speaker is addressing an older person, whereas the Kriol form *jam* may be used in conversation with peer groups or younger people [Meakins 2013: 137].

The table below shows some doublets in the mixed languages addressed here:

Table 7. Doublets in the mixed languages

Meaning	French	Cree	Michif
‘horn’	<i>corne</i>	<i>potachikana</i>	<i>pootachikun, korn</i>
	Kriol	Gurindji	Gurindji Kriol
‘I’	<i>ai</i>	<i>ngayu</i>	<i>ai, ngayu</i>
	Local Quechua	Spanish	Media Lengua
‘dog’	<i>ashku</i>	<i>perro</i>	<i>ashku, pirru</i>

We can propose a hypothesis on the use of doublets in mixed languages. A higher mix in the core lexicon (as in Michif and Gurindji Kriol) correlates with an increase in the amount of doublets. For Gurindji Kriol, we can observe that the core vocabulary comes more or less evenly from both lexifier languages with a significant number of doublets in use. On the other hand, in Medny Aleut with a dominant lexifier language (approximately 80 % of its lexis comes from Aleut), doublets only account

for 7%. An explanation for a larger inventory of doublets may be that with a higher lexical diversity of language, speakers are more likely to use two lexical items from different lexifier languages instead of only one.

However, these statements are very preliminary observations and should be tested in further studies.

8. Discussion and conclusions

In this paper, I considered the core vocabulary in mixed languages to find out whether claims about their vocabulary in general also hold for the stable part of the lexicon. The following languages were examined: Medny Aleut, Michif, Media Lengua and Gurindji Kriol. A look through the lens of the presumably more stable vocabulary (Swadesh 100) has produced the following findings. The core vocabularies of Gurindji Kriol and Michif are more or less equally distributed between their respective lexifier languages, with both showing relatively large proportions of doublets (where a doublet represents two lexical items standing for the same concept). Medny Aleut and Media Lengua, on the other hand, very strongly rely on only one lexifier language, i.e. Aleut and Spanish, respectively, with fewer doublets attested in either of them. The lexification of the core vocabulary in each of the four languages is in line with opinions voiced in the literature on mixed-language vocabularies in general.

This study confirms that in the core vocabulary, just as in their vocabulary at large, only some languages can show an even spread of their lexis between lexifier languages, while in others one lexifier language normally predominates.

Meakins [2018] groups some of the latter, including Medny Aleut, Michif and Gurindji Kriol, under a separate category of ‘structural mixes’. Of these, Medny Aleut, for example, does have a dominant lexifier language, this conclusion also confirmed by an analysis of its less stable vocabulary. On the other hand, Michif and Gurindji Kriol show no strong dominance of one of the lexifier languages over the other, and it proves to have a considerable proportion of doublets.

Based on the above, I hypothesize that mixed languages can show two types of variation. If a language shows a strong variation in terms of where a word for a specific concept may come from historically, then it also shows a strong variation synchronically, with relatively many concepts for which speakers alternatively recruit a word from one or the other lexifier language, depending on the pragmatic context.

Moreover, there is another possible interpretation⁴. According to Auer's model *from code switching via language mixing to fused lects*⁵ [1999], the variation tends to decrease in a diachronic continuum over time. In other words, mixed languages of a more recent origin are less lexically stabilized than older ones. Indeed, as compared to both Michif and Medny Aleut, Gurindji Kriol is of a more recent origin [Meakins 2005] and shows more lexical variation. On the other hand, the origins of Michif are tentatively dated to the early 1800s, [Bakker 1997]. Medny Aleut probably developed in the early 1900s [Golovko, Vakhtin 1990]. Yet, Michif has more doublets than Medny Aleut.

So far, the 'synchronic' hypothesis attributing the number of doublets to the "more mixed nature" of one mixed language as compared to the other seems to be more preferable to the diachronic hypothesis of "incomplete stabilization". Including data from more languages may provide empirical evidence in favor of one or the other of these motivations, or give grounds for yet another interpretation.

References

- Auer 1999 — A. Peter. From code switching via language mixing to fused lects: toward a dynamic typology of bilingual speech. *International journal of bilingualism*. 1999. Vol. 3. Iss. 4. P. 309–332.
- Bakker 1997 — P. Bakker. *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis*. New York: Oxford University Press, 1997.

⁴ I am very grateful for these considerations to the anonymous reviewer.

⁵ Auer used the term 'fused lects' which is very close to the idea of mixed languages (which I use in this article) [Auer 1999: 321]

- Bakker 2000 — P. Bakker. Rapid language change: Creolization, intertwining, convergence. C. Renfrew, A. McMahon, L. Trask (eds.) *Time depth in historical linguistics*. Cambridge: The McDonald Institute, 2000. P. 585–620.
- Bakker 2015 — P. Bakker. Typology of mixed languages. A. Aikhenvald, R. Dixon (eds.). *The Cambridge handbook of linguistic typology*. Cambridge: Cambridge University Press, 2015. P. 217–253.
- Bakker, Matras 2013 — P. Bakker, Y. Matras (eds.). *Contact languages*. Berlin: De Gruyter Mouton, 2013.
- Bakker, Mous 1994 — P. Bakker, M. Mous. *Mixed languages: 15 case studies in language intertwining*. Amsterdam: Institute for Functional Research into Language and Language Use (IFOTT), 1994.
- Bakker, Muysken 1994 — P. Bakker, P. Muysken. Mixed languages and language intertwining. J. Arends, P. Muysken, N. Smith (eds.). *Pidgins and creoles: An introduction*. Amsterdam: John Benjamins Publishing Company, 1994. P. 41–52.
- Belikov 1998 — V. I. Belikov. *Pidzhiny i kreolskie yazyki Okeanii: sotsiolingvističeskij očerok* [Pidgins and Creole languages of Oceania. Sociolinguistic essay]. Moscow: Vostochnaya literatura, RAS, 1998. P. 198–198.
- Belikov 2006 — V. I. Belikov. *Konvergentnye protsessy v lingvogeneze* [Convergence processes in linguogenesis]. Doctoral dissertation. Moscow: Moscow State University, 2006.
- Campbell 2013 — L. Campbell. *Historical linguistics*. Cambridge: MIT Press, 2013.
- Dikker 2008 — S. Dikker. Spanish prepositions in Media Lengua: redefining relexification. T. Stolz, D. Bakker, R. Palomo (eds.) *Hispanisation*. Berlin: De Gruyter Mouton, 2008. P. 121–148.
- Golovko 1994 — E. V. Golovko. Mednij Aleut or Copper Island Aleut: An Aleut-Russian mixed language. P. Bakker, M. Mous (eds.). *Mixed Languages: 15 case studies in languages intertwining*. Amsterdam: IFOTT, 1994. P. 113–121.
- Golovko 2009 — E. V. Golovko. *Aleutian language in the Russian Federation (structure, functioning, contact phenomena)*. Doctoral dissertation. St. Petersburg: Institute for Linguistic Studies, Russian Academy of Sciences, 2009.
- Golovko, Vakhtin 1994 — E. V. Golovko, N. B. Vakhtin. Copper Aleut-English Dictionary. 1994. Manuscript.
- Haspelmath, Tadmor 2009 — M. Haspelmath, U. Tadmor. *Loanwords in the World's Languages: A Comparative Handbook*. Berlin, New York: De Gruyter Mouton, 2009.
- Heine, Kuteva 2005 — B. Heine, T. Kuteva. *Language contact and grammatical change*. Cambridge: Cambridge University Press, 2005.
- Koch, Nordlinger 2014 — H. Koch, R. Nordlinger. *The languages and linguistics of Australia*. Berlin: De Gruyter Mouton, 2014.

- Matras 2000 — Y. Matras. Mixed languages: a functional — communicative approach. *Bilingualism: Language and cognition*. 2000. Vol. 3. Iss. 2. P. 79–99.
- McConvell, Meakins 2005 — P. McConvell, F. Meakins. Gurindji Kriol: A mixed language emerges from code-switching. *Australian journal of linguistics*. 2005. Vol. 25. Iss. 1. P. 9–30.
- Meakins 2013 — F. Meakins. The survey of pidgin and creole languages. S. Michaelis, P. Maurer, M. Haspelmath, M. Huber (eds.). *Gurindji Kriol*. Oxford: Oxford University Press, 2013. P. 131–139.
- Meakins 2018 — F. Meakins. *Mixed languages*. Oxford: Oxford Research Encyclopedia of Linguistics, 2018.
- Mühlhäusler 1981 — P. Mühlhäusler. The development of the category of number in Tok Pisin. P. Muysken (eds.). *Generative studies on creole languages*. Berlin, New York: De Gruyter Mouton, 1981. P. 35–84.
- Muysken 1981 — P. Muysken. Halfway between Quechua and Spanish: The case for relexification. A. Highfield, A. Valdman (eds.). *Historicity and variation in creole studies*. University of Amsterdam, 1981. P. 52–78.
- Myers-Scotton 1998 — C. Myers-Scotton. A way to dusty death: The matrix language turnover hypothesis. L. Grenoble, L. Whaley (eds.). *Endangered Languages*. Cambridge: Cambridge University Press, 1998. P. 289–316.
- Sekerina 1994 — I. Sekerina. Copper Island (Mednyj) Aleut (CIA): A mixed language. *Languages of the World*. 1994. Vol. 8. Iss. 1. P. 14–31.
- Swadesh 1959 — M. Swadesh. Linguistics as an Instrument of Prehistory. *Southwestern Journal of Anthropology*. 1959. Vol. 15. Iss. 1. P. 20–35.
- Thomason 1997 — S. Thomason. *Contact languages: A wider perspective*. Amsterdam: John Benjamins Publishing, 1997.
- Thomason, Kaufman 1992 — S. Thomason, T. Kaufman. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press, 1992.
- Trask 2000 — R. Trask. *Time depth in historical linguistics*. Cambridge: University of Cambridge; McDonald institute for archaeological research, 2000.
- Velupillai 2015 — V. Velupillai. *Pidgins, creoles and mixed languages: An introduction*. (Creole Language Library. Vol. 48). Amsterdam: John Benjamins Publishing Company, 2015.