

IN SEARCH OF MIXED CODES IN THE MEDITERRANEAN: LANGUAGE CONTACTS IN CYPRUS

Petek Kurtböke¹

Gökçin Orundalı

Eastern Mediterranean University, Cyprus

1. Introduction

Cyprus has officially had a Turkish population since it became part of the Ottoman Empire in 1571. After its conquest, Turkish migration began particularly from Southern Anatolia (e.g., Konya, İçel, Antalya, Yozgat, Alanya). Over the centuries, the island had been in the hands of Greeks, Egyptians, Assyrians, Persians, Romans, Arabs, Lusignans and Venetians. In 1878, it was taken over by the British and remained a colony until 1960 when Cyprus became independent. Since 1974, Turkish and Greek communities have been separated by the UN Green Line.

It is difficult to establish the linguistic boundaries of the Cypriot Variety spoken by the Cypriots today, as the migration waves over time have imposed various influences on the language and led to extensive *borrowing* between the populations. The degree of linguistic influence varies according to the length of stay of the conquerors in Cyprus and their interaction with the natives.

While, the majority of loanwords in the Cypriot Turkish and Cypriot Greek (henceforth CT and CG respectively) can be attributed to the mutual influence of these languages on each other, the number of Arabic and Italian loans is also considerable. There remain, however, loanwords, whose origins cannot be identified. This is a common phenomenon in language contact situations and such items and their lexical environment may be taken as an indication of the presence of a mixed code in that particular context.

¹ Petek Kurtböke, P.O. Box 519, Cotton Tree, 4558 Queensland, Australia. Tel/fax.: + 61 7 5443 6695; email: pkurtboke@hotmail.com

2. Mixed code

Researchers have often discussed the formation of a mixed code as an additional choice for communication in language contact situations, and the study of mixed codes has sought answers to a number of questions (Kurtböke, 1998):

- 1) What produces a mixed code?
- 2) What are the linguistic rules which govern it?
- 3) How should such a mixed code be classified?

These questions have not found commonly accepted answers. Commonly, *pidgins and creoles* have been used to clarify the general principles governing the creation of mixed codes (e.g., Mühlhäusler, 1980). Generally speaking, *diachronic approaches* to mixing consider the development of pidgins and creoles. According to Thomason (1995), approaches to language mixing in terms of pidgins and creoles can be identified as follows: mixing without extensive bilingualism and mixing with extensive bilingualism. *Synchronic approaches* to mixing, instead, consider mainly *codeswitching* in a variety of contact situations.

So far, research has been commonly conducted in frameworks which did not allow the mixed code to be studied in its own right but in terms of L1 and L2. However, it is known that in bilingual settings rules of L1 and L2 are not respected and bilinguals often come up with forms not existing in either of the languages. A typical example is Silva-Corvalán's (1995: 6) *en los sábados* 'on the Saturdays' where *en* and *los* should not go together according to either Spanish or English rules. Also Hasselmo (1961) discussed the existence of 3 different codes, 'layers' as he called them, used by the different generations of Swedish migrants in the US. Oksaar (1972), studied Estonian bilinguals in Sweden and in the US and found that her informants violated the rules of both L1 and L2. This occurred particularly in translations and took the form of new collocations. She explained this with a third set of rules developed in addition to L1 and L2. Later, Joshi (1984) developed a computational approach to the explanation of mixed code grammar of Marathi-English code-switching. He observed that the mixed sentences were not always attributable to either L1 or L2, and worked on a computational specification of a set of constraints on the switching rule. Also in the area of Translation Studies, Frawley (1984) observed that the translation itself was essentially a third code arising out of the matrix and target codes, setting its own standards and structure, and it was not always reducible to target and matrix codes.

It has not been clear in the study of Language Contact whether the distinction between language and *code* is purely terminological. The term *code* dates back to the early decades of language contact, and has come to mean that under certain conditions (e.g., immigration), the community language in question would undergo *restriction*. Consequently, when there is reference to mixing in such contexts, the outcome is *code* (e.g., code-alternation, code-switching, code-mixing, code-copying etc.), and it involves *restriction*. And when there is reference to mixing with respect to *pidgins and creoles* the preferred term is *language*, and it involves *development*. The problem with the ‘restriction’ view is that it does not allow the study of what the mixed code might be developing into, although there is consensus that the outcome of language contact may not resemble L1 or L2 structures.

2.1. The mixed code lexicon

The models used in the analysis of loanwords in Language Contact research so far hardly work as they assume that the mixed code lexicon is already full at the start, with the entire vocabularies of L1 and L2. But the problem is to distinguish between what counts as borrowing and what counts as code-switching. Traditionally borrowing has come to be seen as a diachronic process whereas code-switching as a synchronic one, although the criteria for the classification of loans as either one or the other have never been truly reliable. It is unhelpful to force a distinction based solely on frequency criteria and leave aside such important aspects as metaphor (see Kurtböke & Potter, 2000 for a review).

A new approach to the lexicon has been proposed based on corpus research (Sinclair, 1996). In this view, the lexicon is considered empty at the start, contrary to the traditional approach assuming the lexicon full of L1 and L2 vocabularies at the initial stage of contact (Kurtböke, 1998). This does not allow the specification of the mixed code lexicon as an independent construct. This problem can be tackled if initially nothing appears in the lexicon except what is gleaned from the study of the language in use through the study of corpora.

The implications of the empty-lexicon approach for Language Contact research are important. However, a major shift is required in the way the lexicon is perceived. Rather than going to our data with a set of assumptions, we should concentrate on usage. As Sinclair (1996) puts it:

There is no assumption that meaning attaches only to the word; it is anticipated that meanings also arise from the loose and varying co-occurrences of several words, not necessarily next to each other. It is, thus, not possible to compile a list of entries in advance of analysing and interpreting the evidence, because the lexical items are not always words, and each word may enter into a variety of relationships with others to realise lexical items.

Thus, if we examine loans in relation to their environment, rather than as singly-occurring items, and observe their co-selection tendencies, we can see that the meaning of a word and its immediate new context have become inseparable and it may gradually acquire idiom status. Accordingly, we should start building the mixed code lexicon by paying attention to the patterns growing around the “foreign material”.

As for the format of such a lexicon, Sinclair (1996) proposes three components: *a form of a lexical item, an environment, and a meaning*:

A word becomes associated with a meaning through its repeated occurrence in similar contexts. The distinction between the item and its environment is not clear-cut, because the choice of a meaning has a profound effect on the surrounding text, one which is not suddenly cut off at a boundary, but which is correlated with adjacent meanings. Similarly, the domain of meaning does not consist of discrete entities, “meanings”, to each of which can be linked a form; it is assumed to be an amorphous area that is ordered by the number and type of lexical items. Hence the construction of the lexicon requires us to vary all three components against each other [...]. (Sinclair, 1996)

An aspect less considered in language contact is the multiple meanings of words as the “one word-one meaning” attitude to the lexicon prevails. This approach gains importance when we consider the fact that nouns are depicted as the most easily borrowed category. The traditional line of research insists on the centrality of verbs and provides little insight into the behaviour of nouns by treating them as singly-occurring items. Unless we start considering the grammar of nouns in terms of their syntagmatic environment, language contact research will continue along the paradigmatic axis. Syntagmatic relations have received scattered attention in the study of language contact.

A major question facing researchers who favour this approach is ‘*when does the mixed code establish itself?*’ The answer is probably when mixed collocations develop a life of their own. In other words, when *the idiom principle* (Sinclair, 1991) starts operating. This point is related to the debate on grammaticalness vs naturalness as each language contact setting is likely to set its own naturalness criteria. Consequently, emerging units of meaning may not be judged by the criteria of grammaticalness and naturalness of L1 and L2 as they will not resemble either L1 or L2. Within this

framework, it is possible to study the Cypriot Turkish Variety as a mixed code, rather than a dialect of standard Turkish spoken in Turkey, contrary to common practice.

3. Characteristics of the Turkish Cypriot variety

Generally speaking, the Cypriot Turkish Variety is seen as a dialect of the Turkish of Turkey; and the higher the level of education the closer does the speaker draw to the standard language of the mainland (Oakley, 1993). In education, Turkish has been the medium of instruction since the Ottoman times and the Turkish language was officially recognised also under the British rule, although many Turkish-Cypriots, particularly the older generation, speak Greek either learnt at school or picked up in mixed communities. According to 1960 census, some 37% of Turkish-Cypriots were bilingual speakers of Turkish and Greek (Oakley, 1993). The role of bilingual speakers in the diffusion of loanwords across cultures has been a central issue in the study of language contact (Weinreich, 1953; Haugen, 1956; Kurtböke, 1995). Linguistic change is a natural process resulting from such relations between communities.

In fact, Turkish Cypriot Variety differs from the standard mainland Turkish (MT) in all aspects, phonology, syntax, vocabulary and morphology as can be seen in the examples below:

1. “Anan etti börek?” (CT) - Annen börek mi yaptı? (MT) - ‘Has your mother cooked börek?’
2. “Bir kahvecik vereyim?” (CT) - Kahve içer misin? (MT) - ‘Do you want a (quick) cup of coffee?’
3. “Verdi gendine bir elma” (CT) - Ona bir elma verdi (MT) - ‘She gave him an apple’

All three sentences end in rising intonation typical of this variety. These differences between CT and MT, mainly in the spoken language, have been reported to be disappearing since 1974, following a new wave of migration from the mainland (Oakley, 1993).

3.1. Case study: Çatoz

Çatoz is a small village situated 25 miles from Famagusta and 20 miles from Nicosia. It was already a feudal settlement at the time of Lusignans and its name is the Greek version of the French word ‘château’. In 1958, it was given the Turkish name **Serdarlı** in line with the language planning policy of the time. Since the very early

times, the migrant intake has been very low and this has led to the maintenance of the linguistic characteristics of the village. This makes Çatoz linguistically significant and it has been chosen as the pilot area for the current study. The current population is around 1000 with the majority of villagers being 40-60 years of age.

The aim of the study is to create a spoken corpus of samples collected through interviews with a set number of villagers over the age of 40. The interviews of one-hour have been structured around pre-established topics such as folk tales, rituals, festivities, culinary and other practices of local importance. This pilot study is expected to serve as a basis for a larger scale project dealing with the lexicon of the entire Turkish Cypriot Community (Orundalı, in progress).

The Turkish Cypriot variety spoken on the island is commonly referred to as a dialect of mainland Turkish, although the corpus suggests that from a language contact perspective, what we have in hand may be analysed in terms of a mixed code. Some of the borrowings can be immediately recognised as being of Italian origin such as ‘bango’ (banco) and ‘uruba’ (roba), and the Greek influence can be seen particularly in place names as in the preferred village name ‘Kehelaves’ (Keleşler), although not all ‘foreign material’ can be identified as such.

3.2. The use of et-

The use of delexical verb ‘et-’ to nominalize borrowed forms is prominent in the Cypriot Turkish Variety as in other varieties of Turkish in the mainland and in diaspora. This is a lexical unit or a collocation, to use the preferable term in Corpus Linguistics. Delexical verbs in Turkish, namely yap-, et-, and ol- appear frequently in studies of Turkish in contact with other languages (see Kurtböke, 1998 for a review). While, in the mainland Turkish, there seems to be a differentiation in the use of et- and yap- in Cypriot Turkish et- seems more frequent. For example, in Turkey, the singular form of a noun such as *tamir* (repair) attracts *et-* but the plural form *tamirat* (repairs) attracts *yap-* and there is also a difference in meaning. However, in the Cypriot variety, we come across examples such as ‘Anan etti börek?’ (has your mother cooked börek?). This is important information about usage, hence the need for spoken data in such a context. It has been observed that delexicalized verbs are used more frequently in written texts than in spoken ones and in our spoken corpus we find the following non-standard uses of ‘et-’:

- 5) ...‘ne **maraz eden** Usta?’ der... ‘why do you worry, Master? He said’
- 6) ‘bana **pegsemedcik edin**, gidecem’... ‘bake some cookies and I’ll take them with me’
- 7) Padişah Vezire ‘ya söyle ya şimdi seni **cellad ederim**’ demiş. ‘The King said to the Vizir, speak or I’ll have your head cut off’
- 8) Hiç çocuk edmezdi... ‘she couldn’t have any children’
- 9) ...beni babam çoban etmesin mi o tavuklara... ‘My father wanted me to look after the chooks’
- 10) ...kırk gün kırk gece düğün etmişler... ‘the wedding party went on for days...’
- 11) “Mergiz gibi etmişin evi” ‘your house is spotless clean’
- 12) bütün köylü elinden usanmış, giden padişaha labort edermiş... ‘all the peasants were sick and tired of him, they went to the king to complain’

Standard grammars and dictionaries of Turkish do not have very comprehensive treatments of *delexical verbs*. The missing information in traditional grammatical and lexicographic descriptions is the frequent nouns or nominal groups which co-occur with these verbs and carry most of the meaning. Such descriptions are usually limited to the mention of Arabic and Persian verbs borrowed into Turkish as nouns through et- and ol- constructions (e.g., Underhill, 1976). However, these elements facilitate mixing and speed up the nominalisation process under conditions of language contact. Some of the examples above typically belong to ‘borrowed form + delexical verb’ pattern and show the frequency of ‘et-’ in the Turkish Cypriot Variety compared to the frequent delexical verb ‘yap-’ or ‘ol-’ in the standard language (e.g., maraz et-, cellad et-, labort et-). The equivalent of these units in the mainland variety would be as follows:

- 13) ...‘ne **üzülüyorsun** Usta?’ der... ‘why do you worry, Master? He said’
- 14) ‘bana **peksimet yapın**, gideceğim’... ‘bake some cookies and I’ll take them with me’
- 15) Padişah Vezire ‘ya söyle ya şimdi senin **kafanı kestiririm**’ demiş. ‘The King said to the Vizir, speak or I’ll have your head cut off’
- 16) **Hiç çocuğu olmazdı... ‘she couldn’t have any children’**
- 17) **...babam o tavuklara bakmamı istedi... ‘My father wanted me to look after the chooks’**
- 18) **...kırk gün kırk gece düğün yapmışlar... ‘the wedding party went on for days...’**
- 19) “**mis gibi yapmışsın** evi” ‘your house is spotless clean’
- 20) “bütün köylü elinden usanmış, gidip padişaha **söylemişler**” ‘all the peasants were sick and tired of him, they went to the king to complain’

Numerous other examples from Cyprus can be given illustrating the use of *et-* in the place of *yap*, *ol-* and other verbs with delexical function in mainland Turkish. Recent emigration from Turkey to industrialised countries, however, has resulted in an increase in the frequency of *yap-* forms. This difference may be due to a number of reasons including the history of migration waves to Cyprus, how long the conquerors stayed on the island and how their language and culture have influenced the local population.

4. Conclusion

Traditionally, grammars and dictionaries have taken word meanings for granted. However, corpus-driven lexicography has shown that words do not have inherent meanings, but depend on their environments to select and/or confirm their meaning. For example, if a word has two different meanings, a distinctive pattern will grow around each meaning. Consequently, the meaning will have an effect on the environment. Lexicogrammar has major implications for linguistic theory as findings from corpora have challenged long-standing assumptions about lexis-syntax relations. The importance of collocations is increasingly discussed by researchers and the criteria to define these lexical units (from completely frozen to highly variable) are being refined as it is recognized that such units are a central organizing principle in language as well as in the mental lexicon.

During the early decades of language contact research, the lexicon was seen as a simple inventory of items. However, corpus research has revived two Firthian notions, collocation and colligation, as the heart of lexicogrammar. Both notions are useful to explain the inner dynamics of the mixed code which cannot be explained by counting the singly occurring items nicely fitting in a number of grammatical categories. In our study, the non-standard units found in the Cypriot Turkish variety are placed in an empty lexicon, and they are not seen as an extension of Turkish or Greek lexicons or grammars.

Bibliographical references

- Frawley, W (1984). "Prolegomenon to a theory of translation". In W. Frawley (ed.), *Translation*. Ontario: Associated University Presses, 159-75.
- Hasselmo, N. (1961). *American Swedish*. Ph.D Thesis, Harvard University.

- Haugen, E. (1956). *Bilingualism in the Americas. A bibliography and research guide*. American Dialect Society.
- Joshi, A. (1984). "Processing of sentences with intra-sentential code-switching". In D. Dowty et al. (eds.), *Natural language processing: Psycholinguistic, computational and theoretical perspectives*. Cambridge: Cambridge University Press, 190-220.
- Kurtböke, P. (1995). "Bilinguals as the vehicles of borrowing: Sir P Rycout and Lady M W Montagu and some Turkish loanwords in the OED". *The University of Melbourne Working Papers in Linguistics* 15, 48-57.
- Kurtböke, P. (1998). *A corpus-driven study of Turkish-English language contact in Australia*. Ph.D Thesis, Monash University, Melbourne. Retrieval from <<http://home.vicnet.net.au/~petek/thesis/>>
- Kurtböke, P. & L. Potter (2000). "Co-occurrence tendencies of loanwords in corpora". *International Journal of Corpus Linguistics* 5(1), 83-100.
- Mühlhäusler, P. (1980). "Structural expansion and the process of creolization". In A. Valdman & A. Highfield (eds.), *Theoretical orientations in creole studies*. New York: Academic Press, 19-56.
- Oakley, R. (1993). "The Turkish peoples of Cyprus". In M. Bainbridge (ed.), *The Turkic peoples of the world*. London: Kegan Paul International, 85-117.
- Oksaar, E. (1972). "On code switching. An analysis of bilingual norms". *Proceedings of Third International Applied Linguistics Conference* 3, 491-500.
- Orundalı, G. (in progress). *Çatoz Ağzı'nın sözvarlığı üzerine bir çalışma*. M.A. Thesis, Eastern Mediterranean University, Cyprus.
- Silva-Corvalán, C. (1995). "The study of language contact: An overview of the issues". In C. Silva-Corvalán (ed.), *Spanish in four continents studies in language contact and bilingualism*. Washington, D.C.: Georgetown University Press, 3-14.
- Sinclair, J. (1991). *Corpus concordance collocation*. Oxford: Oxford University Press.
- Sinclair, J. (1996). "The empty lexicon". *International Journal of Corpus Linguistics* 1(1), 99-120.
- Thomason, S. (1995). "Language mixture: Ordinary processes extraordinary results". In C. Silva-Corvalán (ed.), *Spanish in four continents studies in language contact and bilingualism*. Washington, D.C.: Georgetown University Press, 15-33.
- Underhill, R. (1976). *Turkish grammar*. Cambridge, Mass.: The MIT Press.
- Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton & Co.