CODE-SWITCHING, LEXICO-GRAMMATICAL FEATURES AND LOAN TRANSLATION: DATA FROM A LARGE MACEDONIAN-ENGLISH CORPUS

Jim Hlavac
Monash University

Abstract. This paper looks at the speech of bilinguals, in this case a group of 103 Macedonian-Australians and focuses on code-switching in a lexical sense, e.g. embedding, insertion, alternation, referring to lexical material that is transferred across languages, regardless of length. Two- or multi-item constructions as a code-switching related phenomenon are investigated, especially the role of light-verb constructions. Lexico-grammatical and lexico-semantic influence (via loan translation) is examined to see whether imported lexical items co-occur with and appear to precipitate structural innovations.

Keywords: code-switching, grammaticalisation, loan translation, light verbs, Macedonian

1. Contact linguistics and code-switching

This paper examines excerpts of speech from 103 bilinguals and focuses firstly on instances of code-switching, here understood as the transfer of lexemes across languages. The statistical incidence and frequency of code-switching is presented, along with examples to illustrate the type of English-origin content-referential items that are found in some people's Australian English-Macedonian speech. Code-switching can be looked at as a phenomenon that refers primarily to the importation of lexical items, and (a high number of) code-switched lexemes alone have historically often been the main

focus of descriptions of language contact situations. Code-switching need not co-occur with or appear to be related to other features of bilingual speech, such as syntactic, phonological or semantic features that may differ from those found in the speech of monolinguals. There may, however, be connections in a statistical and a proximal sense between the presence of other-language lexemes and grammatical structures that bear input from two languages. The phenomenon of inserting other-language items into one's speech, regardless of how habitualised these may be, is unlikely to be restricted to single words only. When conversing with others who share ability in their languages, bilinguals may transfer single words or groups of words, or transfer phrases or collocations as calques (cf. Grosjean's (2001) notion of modes and sociolinguistic features that condition a speaker's level of activation of one or both languages). The latter are also known as 'loan translations' i.e. constructions that bear „any use of morphemes in Language A that is the result of the literal translation of one or more elements in a semantically equivalent expression in Language B" (Backus, Dorleijn 2009: 77). This phenomenon is the focus of the second part of this paper.

The boundary between code-switching of lexemes and loan translation is, however, fluid: constructions that are transferred from another language may be 'completely' rendered via recipient language morphemes, but they also may be only partially rendered, wherein some elements remain 'untranslated'. Such constructions could be labelled 'partial loan translations' or 'lexico-semantic constructions with code-switching'. In regard to the latter, the term 'code-switching' (whether hyphenated or not) has, for many, become a hypernym that encompasses *all* types of cross-linguistic transfer – lexical, semantic, syntactic, phonological, and pragmatic – e.g. „Codeswitching (CS) refers to language use that consists of material from two or more language varieties at any level from the discourse to the clause" (Jake, Myers-Scotton 2009: 207). Other researchers distinguish (lexical) code-switching from the latter, which they label

'semantic transference' (cf. Clyne 2003) or 'selective code copying' of semantic features (cf. Johanson 1998).

As stated, this paper looks firstly at code-switching as a lexical phenomenon and secondly at the incidence of lexico-grammatical and lexico-semantic constructions that may be patterned on models transferred from the other language. These constructions may appear as a sub-set of lexical code-switching where constructions are made up of items from both languages, i.e. a verb phrase in which one or more items may be code-switched across languages, while other elements in the same verb phrase are rendered via recipient language morphemes, though the pattern is source-language influenced. The type of constructions that are most studied are those consisting of a recipient language 'light' verb (hereafter: 'LV') and a source language object. The choice of the LV that is 'pressed into service' is of interest, inasmuch as an existing LV is the one chosen (or another one) and if the LV in such instances appears to be used in a way congruent to its function hitherto. In some cases it appears that a source language collocation is the basis for the form of the partial loan translations. In other cases, a particular type of 'do+OBJ' may assume such regularity that this pattern spreads further, leading to other verbs being used in this way.

My examination of code-switching *and* loan translation (including possible structural innovation) in this same paper does not suggest that I believe that a causative link exists between the two. In contact linguistics, there is the well-known example of speakers of Urdu, Marathi and Kannaḍa in Kupwar, amongst whom code-switching is almost unknown as it is a socially condemned practice, but amongst whom structural change and convergence between the languages of three groups has occurred (Gumperz, Wilson 1971). On the other hand, Friedman's (1994, 1995, 2009) work on Balkan languages shows that along with structural convergence between Macedonian, Albanian, Bulgarian, Greek, Turkish and Romany in particular, there is much evidence that code-switching was also a widespread phenomenon amongst speakers of these languages,

even if a causal link between convergence and code-switching is not overtly made. Myers-Scotton (2002) and Backus (2005) provide more detailed accounts of the relationship that can exist between code-switching and structural change.

Returning to loan translation, sometimes the semantic features transferred from a source-language model result in little more than 'a new expression' that does not violate syntax, as Otheguy (1993: 23) maintains that the phrase [VERB + *para atrás*] in American Spanish varieties still „exploits inherent Spanish possibilities" even though it is probably modelled on the English [VERB + *back*]. Most loan translations may be fixed combinations of two words or morphemes inserted into the recipient language with no structural consequences beyond the distribution of the words or morphemes involved. What can also happen is that the loan translation can take the form of an unconventional combination of morphemes, and with it, a structural innovation occurs. Such an innovation is suggestive of structural change, that is, an on-going process active amongst more than a mere handful of speakers that could, at least theoretically, be ascertained through longitudinal observation. Another thing is, however, important here: a structural innovation such as that mentioned above came about as a largely accidental by-product of loan translation, and the boundary between loan translation and structural innovation may be overlapping.

This paper is descriptive and presents the frequency and incidence of code-switching data. The size of the corpus allows some discussion on propagation, i.e. cross-speaker habitualisation of form. This, in turn, allows speculation on whether what appear as individual occurrences can be conceived of as 'systematically present' features, i.e. to be innovations that are well on the road to becoming examples of *structural change* as a diachronic feature. Lastly, loan translation is looked at, together with the incidence of LV constructions, as a phenomenon associated with structural change. This paper is informed by Myers-Scotton's (2002) approach to language contact phenomena, and one of her recent works on verb

constructions (Myers-Scotton, Jake 2014). Further, this paper draws on Backus and Dorleijn's (2009) and Backus's (2009) work on loan translations and LVs respectively. The legacy of Friedman's work on language change in Macedonian, from both a synchronic and diachronic perspective, is also discernible in this paper.

2. Data sample and profile of informants

This paper is based on the speech of bilinguals whose heritage language, Macedonian, is an immigrant, transposed one in predominantly Anglophone Australia. The 103 bilinguals are mainly first-generation (73), while just over a quarter (27) are Australian-born second-generation speakers. Amongst the first-generation speakers the majority is made up of those born in Aegean Macedonia (northern Greece), who speak a non-standard variety of Macedonian as their L1, who had little or no formal schooling in this language (or in Greek, which only a small number speak as an L2), and who acquired English after emigration to Australia as young adults in the 1950s and 1960s. The other group of first-generation Macedonian-speakers is from today's Republic of Macedonia, and these emigrants received formal instruction in their first language, Macedonian, and sometimes also in English. They emigrated to Australia from the mid-1960s to the mid-1980s.

The second-generation speakers are English-dominant and the variety of Macedonian that they speak is strongly reflective of the vernacular of their parents; only seven received formal instruction in Macedonian in Australia. All informants are residents of Melbourne, a city of 4.2 million people. According to the latest census figures from 2011, approx. 30,000 Melburnians reported that their 'language spoken at home' was Macedonian. A further 30,000 people in Melbourne, mostly second- or third-generation speakers, and those who reported another language as their home language, have proficiency in Macedonian. There is a concentration of Macedonian-speakers across some parts of Melbourne, with two areas having a

concentration level of those born in the Republic of Macedonia of over 10% (ABS 2012).

The data on which this paper is based come from video- or audio-recorded interviews between the informants, either individually or in small groups, conducted not by the present author, but with a fieldworker and data collector, who is himself a second-generation Australian-Macedonian. Informants were usually interviewed in their own homes and the corpus of available data consists of an approx. 10-minute period of each of the interviews that was transcribed by the field-worker[2].

The data sample consists of 95,028 words (or tokens that are lexical items, excluding non-lexicalised forms such as 'uh-huh') and 3244 turns from the 103 informants. The 'unmarked' or predominant language of the recorded interviews was Macedonian. However, as the interviewer himself was an Australian-born, Macedonian-English bilingual known to most informants, the informants were free to code-switch between English and Macedonian. Table 1 below sets out features of the sample relevant to an analysis of code-switching.

There are some differences in the incidence of use of monolingual or bilingual speech between the two generations, and the number of turns that the informants produced that were either monolingual Macedonian, or consisting of both languages (through code-switching), or even monolingual English. On average, 21% of turns contain code-switches. Examples of code-switches are provided with literal glosses and information on grammatical categories. A free translation, together with the following information on the informant are also provided: informant number (e.g. Inf. 25) and generational membership (e.g. Gen.1).

---

[2]   All data were gained through a large-scale project on the language of Macedonian-speakers in Melbourne which included video and audio-taped recordings. Permission to conduct this research on human informants was granted by Monash University Human Research Ethics Committee, Project No. CF10/2366 – 2010001346 on 11 October 2010. A link to a selection of the videos, some of which are sub-titled, can be found at AMHRC (2014).

Table 1: Data on the 76 first-generation and 27 second-generation informants and the number of words, turns and code-switches across the sample.

	First Generation	Second Generation	Total
No. of informants	76	27	103
Ave words per person	951	841	923
Total words	72,297	22,731	95,028
Ave. no. of turns per informant	29	38	31
Total turns	2,215	1,029	3,244
Total Mac. monolingual turns	1,838	667	2,505
Percentage of turns that are Mac. monolingual	83 %	65 %	77 %
Total turns containing code-switches	373	306	679
Percentage of turns containing code-switches	17 %	30 %	21%
Total Eng. monolingual turns	4	56	60
No. of Eng. code-switches / code-switched items	758	708	1,466
Ave. no. of code-switches per informant	11	26	14

The informants' speech is represented here according to its form: the vast majority of the informants are speakers of south-western dialects of Macedonian originating from the Lerin (Gk. Flórina), Kostur (Gk. Kastoriá) and Drama (Gk. Dráma) regions of the provinces of western and central Macedonia in northern Greece (Aegean Macedonia) and from the Bitola region of the Republic of Macedonia. Readers familiar with Macedonian will recognise forms that vary from standard Macedonian[3]. Macedonian is written in the Cyrillic alphabet, but the examples here are given in Roman-script

---

[3]   South-west Macedonian dialects spoken in Aegean Macedonia in the Lerin and Kostur regions have a number of features that distinguish them from Standard Macedonian. Amongst these are: devoicing of some consonants in medial and not only final position - *vapsuvafme* vs. Standard *vapsuvavme* 'dye 1.PL.IMPERF'; loss of consonants in intervocalic position – *ojme vs.* Standard *odime* 'go 1.PL.PRES.'; loss of consonants in final position – *ka:* vs. Standard *kak* 'how'; changes in consonant clusters – *šo* vs. Standard *što* 'what', etc., cf. Friedman (1993) and Vidoeski (2005).

transliterations. See Appendix One for a guide to the transliteration convention employed. The data sample is part of a project on the speech of Australian-Macedonians and aspects of their language use. A sociolinguistic description of this sample of informants is provided in Hlavac (2016).

3. Code-switching

Code-switching data here are presented in general terms to provide a brief insight into the frequency and type of code-switches that are found in the sample. First, a break-up of instances of code-switching is made according to the position of the code-switch within a clause, or at a clause boundary. Examples of all three different types are given below. Table 2 below sets out in statistical terms the frequency of different types of code-switching according to position vis-à-vis clause boundary.

Table 2: Categories and numbers of code-switches
(single-item and multiple item)

	Single items	Multiple items	Total
Extra-clausal code-switching	884	79	963
Intra-clausal code-switching	307	95	412
Inter-clausal code-switching	44	47	91
Total	1235	221	1466

As Table 2 shows, most code-switches (66%) are extra-clausal code-switches, i.e. code-switches that occur at a clause boundary, and which function usually as discourse-specific devices such as *you know* rather than as content-referential items. It is this quality that distinguishes them from inter-clausal switches. In the first place, English-origin items in the Macedonian speech of the informants usually occur in positions where insertion or embedding is morpho-syntactically less constrained, i.e. at clause boundaries. There may also be features about the English items that are inserted

into Macedonian speech to indicate that they may commonly be discourse-pragmatic items that otherwise occur at clause boundaries in English and that their use in Macedonian also coincides with their placement at Macedonian clause boundaries. These items are usually single-item or compound-item code-switches such as *well*, *yeah*, *alright*. Some of these are contained in (1) below:

(1)   **Yeah**, vo, vo Srbija,   ne   to'ku   vo Makedonija.   Po.. pojḱeto
      Yeah, *in, in Serbia,*   NEG *so much in Macedonia.*   *Mo.. more*

      godini vo Srbija gi   pominafme,   **yeah**.
      *years in Serbia   them*ACC.PL *spend*1.PL.IMPERF, *yeah*.

      I   tamu jajca gi   vapsuvafme,   takvi raboti.
      *And   there eggs   them*ACC.PL *dye*1.PL.IMPERF, *such things*.

      Orajt..   i   so   čupinata,   **really…** pojḱe kaj
      Alright.. *and* *with* *girls*+DEF.ART ***really…***   *more at*

      familijata.   **Yeah**, so   familijata   si s..
      family+DEF.ART. ***Yeah***, *with* *family*+DEF.ART REFLEX.PRON

      yeah.. Veligden go   slavevme   sekoja
      yeah.. *Easter*   *it*ACC.SG *celebrate*1.PL.IMPERF *each*

      godina… **yeah, yeah**.
      year…   ***yeah, yeah***.
      Yeah, in, in Serbia, not so much in Macedonia. We spent mo.. more
      *years in Serbia, **yeah**. And we used to dye eggs there, things like that.*
      *Alright.. and with the girls, **really…** more with the family. **Yeah**,*
      *with the family we ourselves, our.. **yeah**. We celebrated Easter every*
      *year… **yeah, yeah**.* (Inf. 89, Gen.1)

In example (1) above, English-origin *yeah* appears six times. Its dictionary-entry function is that of an affirmative which is the function it fulfils at the start of the turn. *Yeah* is, however, poly-functional, here acting as: a pause-filler (second instance), both affirmative and pause-filler (third instance), utterance terminator (fourth instance)

and then turn terminator (fifth and sixth instances). The point is that the function of *yeah* is independent of the morpho-syntactic grid and lexical-referential content of the clauses that otherwise make up example (1) and it is used as a discourse marker with various functions (cf. Hlavac 2006). The two other extra-clausal switches in example (1) above are *orajt* ('alright') as an evaluative or summative marker of the activities that the informant engaged in over Easter, and later, *really* is employed as an amplifier, and possibly also as a pause-filler. Although *really* appears as an adverb that occurs internally in an otherwise Macedonian-language clause, its discourse function is largely separate to the lexical-referential content of the clause, and can therefore be considered to be 'outside' it.

The second-most frequent type of code-switching is intra-clausal code-switching. Example (2) contains instances of this:

(2) ... vo **office** rabotam, um, i   gledam, um, za pulam   za
*... in **office** work1.SG, um, and look1.SG, um, for look1.SG for*

fabrikata   šo   praje. Ah gi   gledam za
factory+DET REL.PRON do3.SG. Ah themACC.PL. look1.SG for

sales šo   praat i   treba   kako, nekako **balance** da
***sales* REL.PRON do3.PL *and need3.SG how, somehow **balance** COMP*

prajime od   **sales** i um   fabrikata,   ama i,   i   dosadno
*do1.PL from **sales** and um factory+DET but and, and   boring*

mi   e.
me.OBJ(IMPERS.SUBJECT)   be3.SG
*Ah, um, I work in, in [an]office, um, and I see, um, I look at how the factory is doing/going.. Ah, I look at the **sales** they are making and I have to like, somehow we have to do a **balance** on the basis of the **sales** and, um, the factory, but it's also boring for me.* (Inf. 83, Gen.2)

In example (2) above, English-origin items that are not phonologically integrated occur in this second-generation speaker's speech, when recounting aspects of his workplace. As stated, most intra-clausal code-switches are single words, and the prominence of nouns,

with a thematic link to the informant's largely English-speaking workplace, is clear. What is also apparent in example (2) is that the English insertions have a lexical-referential function only. The code-switches to English do not have any function at a discourse level as they do not 'do' anything other than refer to English concepts.

The least frequent type of code-switching is inter-clausal. Example (3) contains instances of this:

(3) sega   rabotam … oh, **I work for Aussie Post**, **Australia Post**.
*now   work 1.SG.... oh, **I work for Aussie Post, Australia Post.***

Ah… jas sum   **transport and I'm a van driver**   šo
*Ah... I   be1.SG   **transport and I'm a van driver***   REL.PRON

drive a van and I do, ka   se   veli?..
***drive a van and I do**, how   REFLEX.PRON   say3.SG*

company mail pick-up, so, so ako, ako, ako imaš
***company mail pick-up, so, so**   if,   if,   if   have2.SG*

fabrika i   imaš   pisma, ti   imaš   **contract with**
*factory   and   have2.SG   letters, you have2.SG   **contract with***

Aussie Post i   nie ojme   da   zemame pismata   **and we**
Aussie Post** and we come1.PL COMP take1.PL   letters+DET **and we

take 'em off ya every day i   setne   **we go back to the depot** i
***take 'em off ya every day** and   then   **we go back to the depot** and*

setne toa odi   **out.**
*then   it   go3.SG. **out.***
*Now I work … oh, **I work for Aussie Post, Australia Post**. Ah… I am*
***transport and I'm a van driver** that **drive a van and I do**, how do*
*you say it?.. **company mail pick-up, so, so** if, if, if you have a factory*
*[business] and you have letters, you have [a] **contract with Aussie***
Post** and we come and pick up the letters **and we take 'em off ya
***every day** and then **we go back to the depot** and then that goes **out**.*
(Inf. 63, Gen.2)

There are some intra-clausal switches as well in example (3), but the most prominent feature of this informant's turn are the clause-length descriptions of his duties at work. While English words and morphemes outnumber Macedonian ones in this turn, Macedonian discourse markers frame most of the turn, and conjunctions such as *ako* ('if'), *i* ('and') and the sequential marker *setne* ('then') occur as more macro-level discourse 'signposts' and are supplied mostly from Macedonian. Further, the English verb *drive* remains in its unconjugated 'bare form', although the controlling subject is 3.SG. and the tense PRES.SIMPLE., i.e. one would expect 3.SG. *drives.*

To summarise the main features of the examples above, Example (1) is an instance of code-switching of English forms whose incidence can be accounted for by the pragmatic function that they perform. In fact, it may be possible to speak of something further happening here: the incidence of *orajt* ('alright') and *really* above is characteristic of Australian-English more so than Macedonian pragmatics. This suggests that what may be happening here is not only a replacement of Macedonian discourse markers with English ones, but the adoption of Australian-English pragmatic norms in function *and in form*. I make this observation about pragmatic features here, but do not further look at them as a feature of particular interest in this paper.

In example (2) above, two English insertions, *sales* and *balance*, co-occur with the Macedonian verb *praje*[4] ('to do'). The verb *praje* controls these English items as objects, in the sense of 'make sales' and 'do a balance'. It is of little surprise that items relating to this speaker's work duties are supplied from English as these forms, typically content-referential ones such as nouns, are well documented in other studies on bilingual speech (Clyne 2003, Verschik 2008). In relation to a large Croatian-English bilingual corpus, Hlavac

---

[4]   Macedonian does not have an infinitive, and the 'base' form of a verb (including the one used in dictionary entries) is the 3.SG.PRES. form. The form *praje* here is a non-standard form for standard Macedonian *pravi* 'to do'.

(2003: 93) reported that 686 (71%) of the 962 English single-item code-switches were nouns. Myers-Scotton and Jake (2014: 511) provided a conceptual account of the differential role that nouns play, compared to verbs, such that the former „do not project information about syntactic and argument structure that is included in the abstract structure of finite verbs“. As verbs play a central role in determining relations within the clause through valency and case assignment this role can be an obstacle in facilitating their cross-language transfer in a way that does not apply to other grammatical categories. The choice of verb that co-occurs with an imported English item will be returned to below.

In example (3), the last two words in that example are a Macedonian verb *odi* ('goes') and the English adverb *out*. The meaning of the last two words is 'go out' in the sense of 'leave with no specific destination', as letters and parcels are sent out from a postal despatch centre to various destinations. It appears that this construction is based on the English phrasal verb *go out*. There are equivalent Macedonian constructions available:

(3a) … i   setne   toa   odi   nadvor
     … *and   then   it*NEUT.SG *go*3.SG *outside*
     '… and then it goes outside'

(3b) … i   setne   toa   izleguva
     … *and   then   it*NEUT.SG *leave/exit*3.SG
     '… and then it leaves/exits'

It seems, however, at least in the mental lexicon of this speaker, that none of the possibilities offered by (3a) or (3b) are able to capture his sense of 'goes out'. (It can be presumed these equivalents could have been produced by this speaker as they are sentences commonly heard from second-generation speakers.) In short, the English phrasal verb expression is availed of and appears here with one part of the loan translated, while the other remains untranslated.

4. Loan translation phrases and light verbs

As discussed, nouns, alongside discourse markers, are the most likely candidates for insertional code-switching. In this corpus, over 80% of the 307 single-item intra-clausal are nouns. Amongst those that I allocate to this group are English *-ing* forms, i.e. verb-origin substantivisations. Example (4) below contains an *-ing* form twice, *cycling*, initially in an inter-clausal code-switch, then in an intra-clausal one:

(4) šetam   **with** familijata   i ah   **with friends and also cycling**.
*walk*1.SG.PRES **with** *family*+DET *and ah* **with friends and also cycling**.

sakam   nogu   **cycling** so   **bajsiko.**
*like*1.SG.PRES   *much* **cycling** *with*   **bicycle**+DET
'I walk **with** the family and, ah, **with friends and also cycling**. I like **cycling** a lot by **bicycle**.' (Inf. 85, Gen.2)

Here, *cycling* is referred to as an activity, i.e. as a gerund or verb form that functions as a noun. The second instance of *cycling* follows the Macedonian verb, *sakam* ('I like') and this structure is a transparent *to like* + OBJ one, in which *to like* is a full verb and *cycling* is its object. There are no examples of *cycle* as a code-switched verb attracting Macedonian verbal morphology (e.g. *cyclam* [seɪklam] 1.SG.PRES. 'I cycle'), nor are there congruent analytic constructions with this verb such as *I do cycling*. There are, however, 39 instances of imported *–ing* forms from English, many of which co-occur with the Macedonian verb *praje*, which is employed in such constructions as an LV.

A short digression on *praje* and its use as an LV is warranted here. Firstly, *praje* (and its standard form *pravi*) equates to 'to make' in its primary meaning, and 'to do' in its secondary meaning, e.g. *Taa praje nered* 'She make3.SG mess' = 'She is making a mess' and *Šo praješ?* 'What do2.SG' = 'What are you doing?'. Historically, it was this verb (along with a synonym *čini* 'to do') that performed the function of an LV, often in combination with Turkisms that entered

Macedonian during the Ottoman occupation of the Balkans, e.g. *praje aber* [do3.SG + Tk. *haber* 'news'] meaning 'to inform' (cf. Jašar-Nasteva 1962/1963). Codification of Macedonian in the twentieth century downgraded the LV construction to non-standard or low-register speech. However, the construction remained a part of most speakers' vernaculars, and particularly amongst those in northern Greece, whose Macedonian-speakers remained dislocated from the development of a literary standard[5]. Thus, incidence of *praje* as a grammaticalised DO-verb in this immigrant language contact setting cannot be conclusively attributed to the contact situation (and the influence of the English LV *do*) or to speakers' employment of the most frequent Macedonian verb as an 'attractive' candidate for LV constructions, as there are diachronic data to show it existed previously in homeland varieties of Macedonian. What is of interest for contact linguists is to observe how dispersed the use of *praje* with English *–ing* forms is, and to see if this innovation appears to be widely propagated enough to be considered structural change. Here, *praje* as a grammaticalised LV performs a function in Macedonian similar to that performed by *yapmak* (or *etmek*) in immigrant Turkish (cf. Backus 2009: 307-339) or *tegema* or *saama* as Estonian grammaticalised verbs in Russian-Estonian speech (Verschik 2008: 137, 149).

Across this data sample, there are 515 instances of *praje* (including all its conjugational and tense forms), which makes this by far the most frequent verb in the sample. Further, *praje* co-occurs with 14 of the 39 English *-ing* forms. Below are three examples:

---

[5]   Pontic Greek, spoken by Pontian refugees settled in Aegean Macedonia, also features a DO-verb construction *chtízo* ('I build'), probably modelled on Turkish *yapmak* (Van Hasselt 1972). Those speakers from northern Greece may have had contact with speakers of this variety of Greek, as well as with speakers of Modern Greek, that features compounds with an LV, *kano* ('I do'), such as *kano psonia* ('I do shopping'), alongside *psonizo* ('I shop'). In an émigré setting, *kano,* as an LV with code-switched objects from English, occurs in the speech of Greek-English bilinguals, e.g. *kano jogging* ('I do jogging') and *kano zapping* ('I zap'), (Edwards, Gardner-Chloros 2007: 77).

(5) si   imaja   i   torbite...   luǵeto

REFLEX.PRON. *have*3.PL.PRES *and bags* + DET... *people*+DET

shoppingto ⁶   si   praat.

shopping+DET   REFLEX.PRON. *do*3.PL.PRES

'They also had the bags as well… the people are doing the(ir) shopping.' (Inf. 89, Gen.1)

(6) ...odam   na **gym**... **boxing** praam,   ah i   isto   taka

... *go*1.SG.PRES to **gym**... **boxing** *do*1.SG.PRES. *ah and also   so*

plivam.

*swim*1.SG.PRES

'I go to the gym... I do boxing, ah and I swim as well.' (Inf. 45, Gen.2)

(7) **walking**   praješe?

walking   *do*2.SG.IMPERF

'Did you do (any) walking?' (Inf. 97, Gen.2)

Examples (5) to (7) above show the co-occurrence of *praje* and an English *-ing* form, in both cases as an object of the LV. The question of why this construction occurs needs to be asked. Does it appear that a speaker wishes to employ an English-origin verb form, but does not do so, and instead uses a Macedonian LV, with a substantivised form of the English verb in order to conform to Macedonian syntax? Or is the speaker employing the English *-ing* form as the form that s/he would use anyway in an equivalent English construction – *The people are **doing** their shopping; I **do** boxing; Did you **do** (any) walking?* – and these are loan translations that feature *praje* as a translation of the English LV *do*, which just happens to be the semantic equivalent of the Macedonian LV, *praje*? This question is

---

⁶   The form *shoppingto* 'shopping+DET' (= 'the shopping') is of interest. The phonotactic form and consonantal ending of *shopping* would usually attract the SG.MASC determiner '-ot'. Instead, the speaker employed the SG.NEUT determiner '-to', perhaps influenced by the form of the Macedonian gerund equivalent, *kupuvanje* ('shopping'), which is neuter, yielding *kupuvanjeto* ('kupuvanje+DET').

hard to answer. In response to the first question, in the data there are examples of some English-origin verbs that are morphologically integrated into Macedonian, attracting verb conjugation markers:

(8) celi   tua   se,   izbegaa,   ama pa   se
     *all*MASC.PL *here* *be*3.PL.PRES, *leave*3.PL.AOR *but   then* REFLEX.PRON

     razberame   so   edni,   **ring**uvame
     *understand*1.PL.PRES *with* *some*3.MASC.PL ***ring***1.PL.ITERATIVE.PRES

     pojḱe   so   kanadejci,   tamu   se   moite.
     *more   with   Canadians,   there   be*3.PL *my*MASC.PL.+DET
     'They're all here. They left, but we contact some of them, we ring the Canadians mostly, that's where mine are.' (Inf. 57, Gen.1)

The number of English-origin verbs that occur in the corpus, most of them morphologically integrated into Macedonian, is 11. This is fewer than the 39 *-ing* forms. In general, a non-finite construction to convey English-origin verbs in Macedonian is a preferred strategy (cf. Myers-Scotton, Jake 2014). Those English-origin verbs with Macedonian morphological markers that occur in the sample tend to appear in the speech of older migrants from northern Greece and that of their children, and appear to be habitualised borrowings. This view is based on the author's long-term contact with Macedonian-speakers aside from the forms found in this linguistic sample, e.g. *juzam sé* '*use*1.SG *everything*' = 'I use everything'; *ne se filvam dobro* 'NEG. REFLEX.PRON. *feel*1.SG *well*' = 'I don't feel well'.

In relation to the second question posed above, it is hard to discount the influence of English *do* as a model, as *do* can, alongside its role as a LV with OBJ. *–ing*, now be combined with other common nouns, e.g. 'Let's *do* lunch'; 'I'm *doing* summer'. The distinction between a code-switched English *-ing* form with LV *praje* and what looks like a loan translation is blurred in instances when the English equivalent contains desemanticised *do*. At least one example of *praje* + *-ing* that is clearly modelled on an English construction is the following:

(9) toj   praje   cleaning
     *he   do*3.SG.PRES   *cleaning*
     'He does cleaning' or 'He is a cleaner'. (Inf. 97, Gen.2)

This utterance was made by an informant about another person and the job he performs. A tendency to avoid (perhaps less well known) nouns relating to occupational groups may account for this, such that constructions containing the noun forms such as *toj e čistač* ('he is [a] cleaner') are less common. A similar avoidance amongst other Macedonian-speakers in Australia for nominal terms to refer to occupations was recorded by Stewart (1995: 22), 'popravam karo'i' '*fix*1.SG *cars*' = 'I am a mechanic'; 'rabotam vo banka' '*work*1.SG *in bank*' = 'I am a bank teller'. Loan translation as the causation factor in the following construction is clear where *ima* 'to have' is being employed in a way resembling its use in equivalent English constructions.

Use of *have* in English as a light or heavily desemanticised verb, e.g. 'to have a rest', 'to have dinner', appears to be responsible for the transfer of this role to the Macedonian *ima* in the following example:

(10) … imame   **lunch**
    … *have*1.PL.PRES ***lunch***
    '… we're having lunch.' (Inf. 97, Gen.2)

This use of *ima* as an LV, analogous to the use of English *have,* is unknown in homeland Macedonian. (In homeland Macedonian *jajeme* '*eat*1.PL.PRES.' or *ručame* '*eat lunch*1.PL.PRES.' would be used.) This appears as a nascent grammaticalisation, at least in the idiolect of this speaker, of *ima* as an LV similar to that which is known for *praje*. Stewart (1995: 22) in her corpus from over 20 years ago, recorded a similar example: 'Ќe imam šaoa' 'FUT.MARKER *have*1. SG.PRES *shower*' = 'I will have a shower'.

5. Conclusion

In this paper I have focussed on examples of code-switching in a large bilingual corpus, and presented, in terms of frequency and position vis-à-vis clause boundary, the incidence of different types of code-switching. Extra-clausal code-switches, i.e. affirmatives and high-frequency discourse markers from English make up by far the largest group. This suggests, at least in statistical terms, that pragmatic particles are highly transferable, not only due to their ambient – and therefore perhaps more penetrative – use, but also because they are usually syntactically independent of the grammatical and argument structure projected by other constituents. Intra-clausal code-switching is the second most common form of code-switching found, and within this type of code-switching, nouns are the most prominent group. The content-referential information of an English-origin noun may take the form of a single-item code-switch. The forms that bear other-language content-referential information may be employed as discrete items that have clear-cut boundaries and which have no further effect on the grammatical and semantic structure of utterances other than being lexical insertions. While there may be fewer barriers to the transfer of single forms that are discrete items – as the frequency of discourse markers shows us – it is perhaps counter-intuitive to postulate that the possibility or amenability of transferring content-referential information would be restricted to single lexemes only. Single lexemes may themselves be part of larger constructions.

There are features specific to the language supplying code-switched nouns, here the sub-set of English substantivisations that are *-ing* forms. The development of *-ing* forms is a result of the *-ing* suffix being productive in terms of grammaticalisation in a manner that goes beyond 'continuousness' for verb forms, extending to 'noun-formation', i.e. gerunds, a form with which some verbs *must* collocate (e.g. *to mind* + *-ing*) and a component in other grammaticalisations, such as *to be going to* as FUTURE MARKER [+ INTENTION].

When the English-origin *-ing* (noun) form is inserted into Macedonian speech the lexico-grammatical features of the *-ing* can be 'carried with it', such that this conditions the verb form supplied from the recipient language. In some of the examples provided this frequently coincides with what is also the choice of forms in the equivalent English construction, namely *do + -ing.* It is hard to attribute the incidence of *praje + -ing* in a conclusive way to either the lexico-grammatical features of *-ing* as an OBJ. in LV constructions or to the lexico-semantic features of the *do + -ing,* which is a loan translation from English. Both influences appear to coalesce. Given the frequency of *-ing* forms in English and the prevalence of English as a contributing language in contact situations, it is instructive to look at its incidence in bilingual corpora. Of particular interest are the choice of recipient language forms that surround *-ing* and the possibility that English-origin grammatical and semantic information may co-determine these chosen forms.

Appendix One. Transliteration of Macedonian into Roman-script graphemes.

Macedonian speech is transliterated with the following Roman-script letters, including letters with diacritic marks.

Cyrillic	а	б	в	г	д	ѓ	е	ж	з	ѕ	и	ј	к	л	љ	м
Roman	a	b	v	g	d	ǵ	e	ž	z	dz	i	j	k	l	lj	m

Cyrillic	н	њ	о	п	р	с	т	ќ	у	ф	х	ц	ч	џ	ш
Roman	n	⊠	o	p	r	s	t	ḱ	u	f	h	c	č	dž	š

REFERENCES

ABS [Australian Bureau of Statistics] 2012. Data analysis of 2011 Census. http://www.abs.gov.au/websitedbs/censushome.nsf/home/Data (12.04.2016)

AMHRC [Australian Macedonian Human Rights Committee] 2014. The Macedonian Speech of First and Second Generation Australian Macedonians: A Joint Project between the AMHRC and Dr Jim Hlavac of Monash University http://www.macedonianhr.org.au/contents/163 (12.04.2016)

Backus, Ad 2005. Codeswitching and language change: One thing leads to another – International Journal of Bilingualism 9(3–4), 307-340.

Backus, Ad 2009. Codeswitching as one piece of the puzzle of language change: The case of Turkish *yapmak*. In: Isurin, Ludmila; Winford, Donald; de Bot, Kees. (Eds.) Multidisciplinary Approaches to Code Switching. Amsterdam/Philadephia: John Benjamins. 307-336.

Backus, Ad; Dorleijn, Margreet 2009. Loan translations versus code-switching. In: Bullock, Barbara; Toribio, Almeida Jacqueline. (Eds.) The Cambridge Handbook of Linguistics Code-switching. Cambridge, UK: Cambridge University Press. 75–94.

Clyne, Michael 2003. Dynamics of Language Contact. Cambridge UK: Cambridge University Press.

Edwards, Malcolm; Gardner-Chloros, Penelope 2007. Compound verbs in codeswitching: Bilinguals making do? – International Journal of Bilingualism 11(1), 73–91.

Friedman, Victor 1993. Dialect Variation and Questions of Standardization in Macedonia: Macedonian, Albanian and Romani – Zbornik za filologiju i lingvistiku, 36(2), 7–35.

Friedman, Victor 1994. Turkisms in a comparative Balkan context. In: Septième congrès international d'études du sud-est européen. (Thessalonique, 29 août – 4 septembre 1994) Athens. 521–543.

Friedman, Victor 1995. Persistence and change in Ottoman patterns of code-switching in the Republic of Macedonia: nostalgia, duress and language shift in Southeastern Europe. In: Summer School Code-Switching and Language Contact. Network on Code-Switching and Language Contact. Ljouwert/Leeuwarden, 14–17 September 1994. Fryske Akademy. 58–67.

Friedman, Victor 2009. Turkish infinitives in Balkan Romani: from codeswitching to paradigm shift – Балканско езикознание / Linguistique balkanique XLVIII (1-2), 27–32.

Grosjean, François 2001. The bilingual's language modes. In: Nicol, Janet. One mind, two languages: Bilingual language processing. Oxford and Malden, MA: Blackwell. 1–22.

Gumperz, John; Wilson John 1971. Convergence and creolization: a case from the Indo-Aryan/Dravidian border in India. In: Hymes, Dell (Ed.) Pidginization and creolization of languages. Cambridge, UK: Cambridge University Press, 151–167.

Hlavac, Jim 2003 Second-generation speech. Lexicon, code-switching and morpho-syntax of Croatian-English bilinguals. Bern: Peter Lang.

Hlavac, Jim 2006 Bilingual discourse markers: Evidence from Croatian-English code-switching – Journal of Pragmatics 38, 1870–1900.

Hlavac, Jim 2016. Three generations, two countries of origin, one speech community: Australian-Macedonians and their language(s). Munich: Verlag Otto Sagner.

Jake, Janice; Myers-Scotton, Carol 2009. Which language? Participation potentials across lexical categories in codeswitching. In: Isurin, Ludmila; Winford, Donald; de Bot, Kees. (Eds.) Multidisciplinary Approaches to Code Switching. Amsterdam/Philadephia: John Benjamins. 307–336.

Jašar-Nasteva, Olivera 1962/1963 Makedonski kalki od turskiot jazik – Makedonski Jazik 13-14, 109–172.

Johanson, Lars 1998. Frame-changing code-copying in immigrant varieties. In: Extra, Guus & Verhoeven, Ludo (Eds.) Bilingualism and migration. Berlin: Mouton de Gruyter. 247–260.

Myers-Scotton, Carol 2002 Contact Linguistics. Bilingual Encounters and Grammatical Outcomes. Oxford: Oxford University Press.

Myers-Scotton, Carol; Jake, Janice 2014. Nonfinite verbs and negotiating bilingualism in codeswitching: Implications for a language production model – Bilingualism: Language and Cognition 17(3), 511–525.

Otheguy Ricardo 1993. A reconsideration of the notion of loan translation in the analysis of U.S. Spanish In: Roca, Ana; Lipski, John (Eds.) Spanish in the United States: Linguistic contact and diversity. Berlin/New York: Mouton de Gruyter, 21–45.

Stewart, Betty 1995. Some observations on the Speech of Bilingual Macedo-
    nians. In: Bivell, Victor (Ed.) Macedonian Agenda. Sydney: Pollitecon.
    15–26.

Van Hasselt, F. G. 1972. Idiomatic parallels in Turkish and Modern Greek.
    In: Bakker, Willem; van Gemert Arnold; Aerts, Willem (Eds.) Studia
    Byzantina et Neohellenica Neerlandica Leiden: Brill. 270–280.

Verschik, Anna 2008. Emerging Bilingual Speech. From Monolingualism to
    Code-Copying. London/New York: Continuum.

Vidoeski, Božidar 2005. Dialects of Macedonian [Translated by Paul Foster].
    Bloomington, IND: Slavica.