Definite article switches in the speech of Bulgarian/Greek bilinguals

Emma Mileva
Simon Fraser University

1 Introduction

1.1 The Balkan Sprachbund

The term Sprachbund (Trubetzkoy 1928), or unité linguistique (Sandfeld 1930), refers to a situation where linguistic communities, which are not necessarily genetically related, are in geographical proximity and share a number of linguistic features due to contact. The Balkan languages included in the Sprachbund are all Indo-European (excluding Balkan Turkic), but they belong to different language branches: Albanian, Modern Greek, the South-Slavic languages Bulgarian, Macedonian and Serbo-Croatian, the Romance languages Daco-Romanian, Aromanian (Vlah) and Judeo-Spanish (Ladino), the Indic language Romani (the language spoken by the Roma in the Balkans), and Turkish from the Altaic family (Joseph 2003). Some authors consider as core participants only Albanian, Modern Greek, the South Slavic languages and Romanian, and all the rest as marginally related (Asenova 2002, Winford 2003). It should also be noted, that "Balkan languages" is not equivalent to "languages of the Balkans". The latter comprises other languages such as Armenian, German, Ukrainian, Yiddish, etc. which are spoken in the Balkan Peninsula, but don't share the structural similarities of the former and thus, are not included in the Balkan Sprachbund.

Linguistic areas emerge as contact between the various speech communities is enhanced by numerous historical events involving periods of conquest, war, colonization, etc. In the case of the Balkan area, these events date back between 800 and 1,700 AD (Winford 2003). Such dramatic influences aside, language contact and change are enhanced if enough users alter their speech based on peaceful economic forces, such as worker migration to more prosperous neighboring countries. As Joseph (1983) points out, it is equally likely that mutual accommodation and shift among the immigrant groups themselves promoted the spread of features.

What makes such linguistic areas complex and intriguing to linguists is the great degree of convergence at several levels of linguistic structure—phonological, morphological, lexical, and syntactic. The intriguing part is that these abstract similarities cannot be explained genetically because the shared features, or arealisms (in this case balkanisms), and do not represent common inheritances from Proto-Indo-European (Joseph 2003). They are the result of linguistic convergence resulting from prolonged and intense contact between the different speech communities. The most discussed Balkan features are: a simplified nominal case system, the merging of dative and genitive cases, the

formation of a future tense using reduced form of the verb 'want', use of postposed definite article, use of evidential, loss of the infinitive, analytical adjectival comparative structures, and clitic doubling.

1.2 Postposed definite article

One of the most discussed topics among Balkan linguists is that of grammaticalized definiteness. Definiteness is a feature of noun phrases distinguishing between entities that are specific and identifiable in a given context and those that are not. One of the ways to express definiteness is to use the definite article. It can be introduced either from a previous discourse context, or accessible through the speech situation through general knowledge, or via a relation that has been established to a separate identifiable referent (a nominal or a relative clause) (Matras 2003).

As for their morphological appearance they can be postposed (in Bulgarian, Macedonian, Albanian, and Romanian) or preposed (in Greek). Historically, the Bulgarian definite article has its origin from initially demonstrative pronouns that later acquired a meaning and function of a definite article. In Greek the definite article is placed before the nominal: *o anthropos* 'the man' and was based, too, on native material: the pronoun that became an article was mostly demonstrative in the earliest periods of the history of the language (Mirchev 1978).

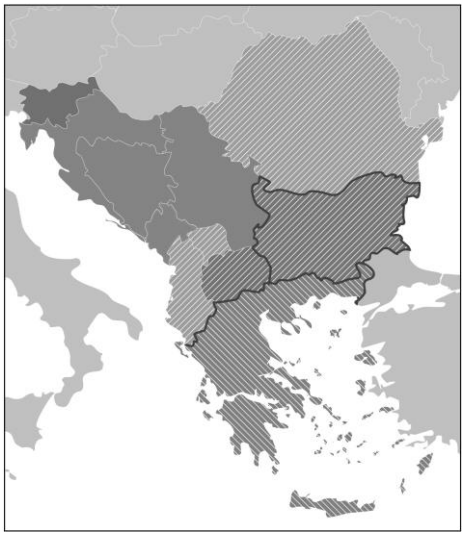

Figure 1. Map of distribution of preposed and postposed definite article in the Balkans. Postposed: forward slanting lines; preposed: backward slanting lines (Perry-Castañeda Library Map Collection, University of Texas; additionally designed).

1.3 Code-switching

Code-switching is a frequently observed phenomenon in the speech of bilingual communities around the world (Backus 2003, Auer 1995, Milroy and Muysken 1995). As Myers-Scotton (2002) notes, code-switching is a common phenomenon in linguistic areas such as the Balkans. It occurs among multilingual speakers, and

broadly refers to the mixing of two or more languages in discourse. Increased contact leads to multilingualism, multilinguals code-switch and as code-switching becomes more intense structural features can pass from one language to another. Vogt (1954) suggested that language contact, including language alternation, is an important element of language change and is a major vehicle for convergence (Gardner-Chloros 2004).

Code-switching allows for alternating between complete utterances from the two languages (inter-sentential code-switching) or between structures related to the sentence or clause, and insertion of lexical items from one language to another (intra-sentential code-switching). This intra-sentential code-switching creates utterances with lexicon and morphosyntactic structure from one language and insertion of a single word or phrases from the other. As a result three kinds of constituents can be produced: mixed constituents consist of materials from both languages; embedded language islands are phrases incorporated from the embedded language; and matrix language islands coming entirely form the matrix language.

Recent research on intra-sentential code-switching investigates the ways that the integration of the embedded language units is accomplished morphosyntactically. Following Joshi's (1985) Closed Class Items Constraint, Myers-Scotton (1993) suggests a model for looking at code-switching data based on a language production process.

The underlying idea of the Matrix Language Framework (Myers-Scotton 1993) is that in code-switching interaction, one of the languages has a dominant role defined as a Matrix Language, while the other is less active—the Embedded Language.

Two major principles are operating in the code-switching processes, the morpheme order principle and the system morpheme principle. Both of these come from the same language, which should be determined as the Matrix Language. In other words, the Matrix Language gives the morphosyntactic frame of the code-switched sentences by providing the surface morpheme order and productive system morphemes (productive inflections and function words).

Another cornerstone in the Matrix Language Framework is the distinction between system and content morphemes. Psycholinguistically, this demarcation is based on their different behavior in the mental lexicon of the speakers. They are activated on different levels of a language production process. The criteria for distinguishing them are [+/−Quantification], [+/−Thematic Role-Assigner] and [+/−Thematic Role-Receiver]. A system morpheme is "any lexical item or affix that is a member of a syntactic category which involves quantification across variables and thus shows [+Quantification] property," while content morphemes are "any categories which show the [−Quantification] property but have either the

[+Thematic Role-Assigner] or [+Thematic Role-Receiver] property" (Myers-Scotton 1993:100).

Stemming from this classification quantifiers, possessives, determiners, dummy pronouns, tense/aspect, complementizers, agreement markers, copula, 'do' verb, possessive 'of', degree adverbs are system morphemes. The content morphemes include nouns, verbs, adjectives. The occurrence of morphemes in Matrix Language + Embedded Language constituents is determined by their status. Under the Matrix Language Framework all the system morphemes should come from the Matrix Language but content morphemes can be from both participating languages.

Another important issue in this framework, and in the code-switching literature, in general, is the domain of intra-sentential code-switching. For the purposes of this study, I follow Myers-Scotton's view that the best unit of analysis for examining code-switched data (and in general for any contact phenomena) is the CP or the maximal projection of complementizer (Myers-Scotton 2002:53).

The objective of this study is to examine instances of single-morpheme switches where the morpheme is a definite article combined with a Greek noun within a Bulgarian clause. The data show that speakers have three choices: they can use the Greek article (1a), the Bulgarian article (1b), or both (2). I explore which factors favor each choice, and what they tell us about the nature of code-switching.

(1a)	*to*	*mathima*		(1b)	*ðilima*	*-ta*
	GR DEF.ART.N.SG	GR NOUN	**vs.**		GR NOUN	BG DEF.ART.N.SG
	'the lesson'				'the dilemma'	

(2)	*Ama*	*do*	*leshi-to*	*ne*	*stigna*	*li?*
	but	to	canteen+DEF.ART.N.SG	NEG	reached.2SG	INT PRT
	'What, didn't you make it to the canteen?'					

2 Methodology

The study was conducted in Thessaloniki, Greece and involved eight adult participants (four males and four females) between the ages of 25 and 30. Their native language is Bulgarian and they have been living in Greece for more than three years, some of them even more. They also possess a high level of bilingualism, and live in a social environment that requires continuous shifts from one language to another. These subjects were recruited by way of the 'friend of a friend' approach (Tagliamonte 2006). Six of them have been learning Greek intensively in their native country, Bulgaria, since the age of 18, and one female

and one male since their early childhood. All eight participants knew each other through common friends, six of them studying in the same department at the university.

The participants were involved in digitally-recorded (Marantz recorder, model No. PMD 660) one-hour interviews in an informal setting without being aware of the study's ultimate objectives. The interviews took place either at the researcher's hotel or the participants' homes, sometimes involving a pair of participants.

The elicitation type of questions I asked were related to their studies in Greece, their way of living and the type of difficulties they had to overcome in order to adapt to their new environment. Other questions inquired whether they considered staying in Greece after graduation, the cultural differences and similarities between the two countries, and how they made friends there.

Apart from these common topics, the talks were shaped according to each interviewee's interests and background. Often, I asked more community-specific questions (Tagliamonte 2006), or more tailor-made ones, related to their personal experiences such as: "Tell me something more about your research work last summer on the islands?" or "What happened when you found out your exam had been marked with an unreasonably low grade?"

The conversation often took different directions from the ones I wanted to follow but that actually was beneficial because it made the whole experience much more spontaneous, without me interrupting the interviewee to purposefully shift the direction. My role here was more as an observer, than interviewer, because I was no longer asking but just passively participating in the discussions and letting them unfold. Turn-taking between different participants made the conversation lively. The more relevant and "controversial" the topic was to them, the more they conversed and enriched the discourse.

In evaluating my data, I would say that the participants code-switched more when the recorder was off. This fact comes to confirm that the Observer paradox plays a crucial role in gathering data in a sociolinguistic interview. It was maybe the main obstacle that limited the amount of code-switched items in my data compared to the other challenges I faced.

The topic was a factor also that influenced the degree of code-switching. Overall, it was observed that code-switching occurred mostly and with all participants when they talked about their way of living in Greece and their problems related to that. This is the case also with topics related to their student life, both about the university and leisure. All of them showed code-switching when telling funny stories or jokes, retelling movie plots, or something they had recently watched on television. Not so much code-switching occurred when they talked about friends, places or events that had happened in Bulgaria.

Overall, the interviews can be characterized as casual in terms of style and despite the concerns listed above they can be defined as naturally-occurring conversations. The purpose of this study was to obtain enough data for analyzing structural aspects of code-switching and not for social ones. In this respect, the collected data were appropriate and suitable.

3 Data analysis and results

The speech of the participants was transcribed using Praat software, and then transferred to Word, Excel and GoldVarb software during the data analysis stage. From the eight and a half hours of transcribed speech only the CPs including code-switching served as the base for the analysis dataset. Approximately 1% of these were indistinct tokens due to overlapped speech and were therefore excluded.

The choice between the two types of construction in Table 1 can be treated as a linguistic variable. In order to explore the variation, the following coding schema with three factors groups was designed:

Factor Group	Factors
FG1 Greek noun ending	[a], [i], [o], C
FG2 Preceding word	B = Bg lexical item G = Gr lexical item
FG3 Subject - Object	S = subject O = object

Table 1. Coding schema for the variants.

The overall number of tokens was too low to conduct a logistical regression. Nevertheless, the distribution of the variants in the different environments still allows us to make some salient observations. In Table 2 the distribution of the two variants across factor groups are represented.

Factor Group (N=60)	VARIANT				TOTAL	
	BG definite article with Greek noun		GR definite article with Greek noun			
	Tokens	%	Tokens	%	Tokens	%
Greek noun ending						
[a]	22	73.3	8	26.7	30	50.0
C (consonant)	10	71.4	4	28.6	14	23.3
[o]	7	100.0			7	11.7
[i]	5	55.5	4	44.5	9	15.0
Preceding lexical item						
Bulgarian	37	84.1	7	15.9	44	73.3
Greek			10	100.0	10	16.7
no presence	4	66.6	2	33.4	6	10.0
Syntactic function of the Greek noun						
Subject	13	72.2	5	27.8	18	30.0
Object	31	73.8	11	26.2	42	70.0

Table 2. Distribution of variants across factor groups.

The overall proportion in percentage between the two variants is 73% (Bulgarian definite article) vs. 27% (Greek definite article). Preceding Lexical Item appears to be the most significant factor group compared to the others. Bulgarian preceding word is found in 44 tokens. In more than half of these (about 62%) the grammatical category of the item is a preposition.

Looking at the Matrix Language of the sentence where these tokens are found in the data, it can be easily seen that this is Bulgarian, and the Greek noun is considered as Embedded Language. A Greek word as a preceding element does not occur at all. The Greek definite article is used with just 7 tokens when the preceding element is a Bulgarian word, and in 10, when it is a Greek word. The factor group Greek Noun Ending does not appear to be a significant factor. Yet it is interesting to see that from a total 30 Greek nouns in the data ending in –[a], 22 of them were assigned the Bulgarian definite article for feminine singular or plural. The syntactic factor group Subject–Object apparently does not determine the speaker's choice. Rather it seems that Greek nouns with Bulgarian articles occur far more frequently (31 tokens) as objects in the clause, than with Greek articles (13 tokens). In the position of a subject, preferences are for using Greek definite article more than Bulgarian one.

An interesting finding in the data is related to Greek proper names combined with Bulgarian definite article. Proper names in Greek as in Albanian, and in some cases in Turkish (Napoli 2009:588) are obligatory accompanied by a

definite article. In contrast, in Bulgarian (and Romanian) proper names do not take a definite marker, excluding isolated cases where definite articles are added to diminutives. However, the data shows that proper Greek names are used with a Bulgarian definite article:

(3)	Nali	kazvali,	che	Mesoghio-to	i
	INT PRT	say. 3PL.PPAA	that	Mediterranean+DEF.ART.N.SG	and
	Egheo-to	bili	naj-mrysnite	moreta	
	Aegean+	be.3PL.PPAA	SPRLdirty	seas	
	DEF.ART.N.SG				

'They said that the Mediterranean and Aegean seas were among the most polluted seas.'

If a definite article serves as an indicator of non-ambiguity of reference (Christophersen 1939) why do the speakers use it with objects definitely unique in the world? There is only one Mediterranean Sea in our physical world and the Aegean Sea experiences the same semantic characteristic of uniqueness.

4 Conclusion

By using an ethnographic approach and interview techniques this study aimed to explore code-switching patterns with focus on definite article morpheme switches. Eight Bulgarian-Greek bilinguals were interviewed in informal settings and the data was analyzed using variationist approach. Even though the collected data did not give enough tokens for statistical analysis, some interesting observations were made.

The data provides support for Myers-Scotton's Matrix Language Framework (Myers-Scotton 2002) in terms of the distinction between system and content morphemes. Looking at the sequence of cases with Bulgarian preposition—Greek noun—Bulgarian definite article the schematic representation is as follows:

Preposition	Noun	Definite article
BG	GR	BG
System	Content	System

Table 3. Sequence of system and content morphemes.

This supports her theory at least with respect to classifying definite articles and prepositions (albeit not all) as system morphemes, i.e. they neither assign nor receive thematic roles, and also supports her prediction that they should come

from the Matrix Language. According to Matrix Language Framework, system morphemes form a 'nest' for the insertion of content morphemes, and my data support that prediction. As to the type of prepositions, most frequently occurring are those with most abstract and general relationships in Bulgarian (Asenova 2002): *v* 'in', *na* 'of, on', and *za* 'about, for'. It was also shown that preferences for using a Bulgarian definite article with Greek noun are not phonologically motivated but dependant on the preceding lexical item.

It was observed that to proper Greek names a Bulgarian definite article was added. This poses a question whether this is a nascent structural (and semantic) transfer from one language to the other due contact or is it purely coincidental?

And lastly, the switches between a lexical stem and a bound morpheme as in the case of a Greek noun and a Bulgarian article are additional counterexamples against Poplack's (1980) notion of the Free morpheme constraint. It is one of the early attempts to determine allowed and prohibited mechanisms of code-switching, which states that only free, and not bound morphemes, can be switched in bilingual utterances. An exception is the cases when the stem has been 'phonologically integrated into the matrix language of the morpheme' (Winford 2003:128). Further support for this violation is provided by Eliasson (1995) on Maori-English, Zabrodskaya (2008) on Estonian-Russian, and Myers-Scotton (1993) on Swahili-English, among others.

References

Asenova, P. 2002. *Balkansko Ezikoznanije*. Veliko Tarnovo: Faber.

Auer, P. 1995. The pragmatics of code-switching: A sequential approach. In *One speaker, two Languages: Cross-disciplinary Perspectives on Code-switching*, ed. by L. Milroy and P. Muysken, 115–135. Cambridge: Cambridge University Press.

Backus, A. 2003. Units in code switching: Evidence for multimorphemic elements in the lexicon. *Lingusitics* 41.83–132.

Christoferson, P. 1939. The articles. A study of their theory and use in English. Oxford: Munksgaard, Copenhagen and Miford.

Eliasson, S. 1995. Grammatical and lexical switching in Maori-English "grashopper speech". Summer School on Code-switching and Language Contact. Ljouwert/Leeuwarden, The Netherlands, 14–17 September, 1994.

Gardner-Chloros, P. 1995. Code-switching in community, regional and national repertoires: The myth of the discreteness of linguistic systems. In *One speaker, two Languages: Cross-disciplinary Perspectives on Code-switching*, ed. by L. Milroy and P. Muysken, 68 – 90. Cambridge: University Press

Gardner-Chloros, P. 2004. How code-switching mediates politeness: Gender-related speech among London Greek-Cypriots (with Katerina Finnis). *Estudios de Sociolinguistica* 4.505–533.

Joseph, B. 1983. Language use in the Balkans: The contributions of historical linguistics. *Anthropological Linguistics* 25.275–287.

Joseph, B. 2003. The Balkan languages. In *International Encyclopedia of Linguistics, 2nd edition*, ed. by W. Frawley, 153–155. Oxford: Oxford University Press.

Joshi, A. 1985. Processing sentences with intra-sentential code switching. In *Natural Language Parsing*, ed. by D. Dowty, L. Karttunen, and A. Zwicky, 190-205. New York: Academic Press.

Labov, W. 1984. Field methods of the project on linguistic change and variation. In *Language in Use: Readings in Sociolinguistics*, ed. by J. Baugh and J. Sherzer, 28–54. Engelwood Cliffs, NJ: Prentice-Hall.

Milroy, L., and P. Muysken. 1995. One Speaker, two Languages. Cross-disciplinary Perspectives on Code-switching. Cambridge: Cambridge University Press.

Mirchev, K. 1978. *Istoricheska Gramatika na BalgarskiaEezik.* Sofia: Nauka i izkustvo.

Myers-Scotton, C. 1993. Dueling languages: Grammatical Structure in Code-switching. Oxford: Clarendon Press.

Myers-Scotton, C. 2002. *Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University.

Napoli, M. 2009. Aspects of definiteness in Greek. *Studies in Language* 33.569–611.

Poplack, S. 1980. "Sometimes I'll start a sentence in English Y TERMINO EN ESPANOL": Towards a typology of code-switching. *Linguistics* 18.581–618.

Sandfeld, K. 1930. *Linguistique balkanique. Problèmes et Résultats*. Paris: Librairie ancienne Honoré Champion.

Tagliamonte, S. 2006. *Analyzing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Troubetzkoy, N. 1928. Studies in general linguistics and language structure. In *First International Congress of Linguistics, Actes, 17f*. Leiden. Proposal 16.

Vogt, H. 1954. Language contacts. *Word* 10.365–374.

Winford, D. 2003. *An Introduction to Contact Linguistics.* Oxford: Blackwell Publishing.

Zabrodskaja, A. 2008. Code-switching as a mechanism of contact-induced language change in Estonia's Russian. Conference on Cognitive and Functional Perspectives on Dynamic Tendencies in Languages, University of Tartu, Estonia.