THE BALKAN SPRACHBUND

Linguistic Institute 2017
Brian D. Joseph

Balkans

Neighboring Languages with shared linguistic traits:

Albanian

Slavic:
  Bulgarian
  Macedonian
  some parts (mostly SE Serbian, the Torlak region) of the Bosnian-Serbian-Croatian-Montenegrin complex)

Italic (Romance):
  Aromanian
  Megleno-Romanian
  Daco-Romanian
  Judezmo (to a lesser extent)

Romani

Turkish

"Sprachbund":

= a linguistic area where languages, through intense and
    sustained contact in a <mark>mutually multi-laterally multi-
    directionally multilingual</mark> society, have come to converge
    with one another structurally and lexically and to diverge
    from the form that they held previously

Cultural area too:

pan-Balkan foods and drinks:

Glasses of *tsipouro / raki …*

Pitas (*spanokotiropites / börek / banitsa / maznik …*)

Kebabs (*souvlaki / qebap / kebap / …*)

Roasted red peppers

Back to language:

J. Kopitar (1829):

"*nur eine* Sprach*form* herrscht, aber mit dreyerley Sprachmaterie"

"only one grammar ('language-form') dominates but with three lexicons ('threefold language-material')"

i.e., as a slogan: "Three languages, one grammar"

For instance:

• in the form of the "verbal complex" (verb + modifying affixes and grammatical markers that can stand alone as a sentence); thus, in the following, meaning 'I will not give it to him', one finds all the languages with literally "not.will.that.to-him.it.give-I":

(Albanian)

    a. *s'*   *do*   *të*   *j*   *a-*   *jep*
        NEG   FUT   SUBJVE   him/IO   it/DO   give/1SG

(dialectal Greek)

    b. *ðe*   *θe*   *na*   *tu*   *to*   *ðóso*
        NEG   FUT   SUBJVE   him/IO   it/DO   give/1SG

(Daco-Romanian)

	c.	*nu*	*o*	*să*	*i*	*-l*	*dau*
		NEG	FUT	SUBJVE	him/IO	it/DO	give/1SG

(dialectal Macedonian)

	d.	*ne*	*ḱe*	*da*	*mu*	*go*	*davam*
		NEG	FUT	SUBJVE	him/IO	it/DO	give/1SG

• for 'I feel like VERB-ing':

Thessalian Grk: *mi*   *trogiti*
                      to-me it-is-eaten

Albanian   *më*   *hahet*
                      to-me it-is-eaten"

Macedonian   *mi*   *se*   *jade*
                      to-me REFL eat

(NB: Slavic reflexive structure is functional equivalent of Greek/Albanian passive structure

• **for 'y'all give me!' (i.e., "give/PL (to-)me"):**

Thessalian Greek: *ðo-m-ti*
'give-me-2PL!'

Albanian: *jep-e-ni*
'give-it-2PL!'

(both with an object pronoun inside of PLURAL marker)

MORE ON THE LANGUAGE BACKGROUND

Languages of the Balkans (= on a geographic basis)

ANCIENT (not all of equal antiquity):

Dacian (Daco-Mysian) [= pre-Romanian??]

Greek

Illyrian [= pre-Albanian??]

Macedonian [??sibling to (all of) Ancient Greek or an Ancient Greek dialect??]

Old Church Slavonic

Phrygian

Pre-Greek ("Pelasgian")

Thracian [= pre-Albanian????]

+ traces of Continental Celtic and some Germanic (Gothic and/or Norse); possibly
   also an Armenian presence prehistorically

MODERN [not counting very recent migrations, e.g. by Arabic or Tagalog speakers into Greece]

Albanian

Armenian (spoken in Bulgaria)

Bulgarian

Circassian (Adygey variety; spoken in parts of Kosovo (in former Yugoslavia)

German (spoken in Romania)

Greek (including the very divergent dialects like Tsakonian and Pontic (the latter only in Balkans proper via relatively recent migrations))

Hungarian (spoken in Romania)

Italian (spoken in Istria area of (former) Yugoslavia)

Judezmo (also known as Ladino or Judeo-Espagnol)

Macedonian [NB: ≠ Ancient Macedonian]

Romanian (fuller picture to come, below)

Romany [= Indic language of the Gypsies]

Ruthenian (also known as Rusyn, spoken in Vojvodina area of (former) Yugoslavia (= a dialect of Ukrainian??))

"Serbo-Croatian" [= (now) Bosnian-Croatian-Serbian (??Montengrin??)]

Slovak (in a small enclave in Vojvodina area of (former) Yugoslavia)

Slovenian

Turkish

Balkan Languages (= those participating to some significant extent in Balkan Sprachbund, less-involved languages/dialects in *italics*, though degree of involvement varies considerably)

Albanian (major dialects: *Geg* **(North) and Tosk (South))**

Bulgarian

Greek (various dialects, including Tsakonian (but excluding Asia Minor dialects))

Judezmo **(maybe only at phonological (and to some extent lexical) level at best)**

Macedonian

Romanian (actually more specifically Aroumanian (Vlach), and Megleno-Romanian, less so *Daco-Romanian* **and even less so** *Istro-Romanian***)**

Romany (= Indic language of the Gypsies)

Serbian **(really only via Torlak dialects of Southeast as most relevant; much less so** *Croatian***,** *Bosnian***, etc.)**

Turkish **(more or less, not a "full" structural participant but crucial nonetheless)**

Balkan convergent features ("Balkanisms")

Those most discussed in the literature, covering morphosyntax (a - g), semantics/pragmatics (h), syntax (i - j), lexical derivation (k), and phonology (l - q):

Distribution of features (* = partial or dialectal realization; "Slavic" = general across Balkan Slavic; "Romance" = general across Balkan Romance:

a. a reduction in the nominal case system, especially a falling together of genitive and dative cases [Alb, Grk, Slavic, Romance]

b. the formation of a future tense based on a reduced, often invariant, form of the verb 'want' (not unlike Eng *I will write*) [Alb*, Grk, Slavic, Romance*, Romany]

c. the use of an enclitic (postposed) definite article, typically occurring after the first word in the noun phrase (thus "man-the good" for 'the good man') [Alb, Slavic, Romance]

d. analytic adjectival comparative adjective formations (thus "more strong" for "stronger") [Alb, Grk, Slavic, Romance, Romany, Trk]

e. marking of personal direct objects with a preposition (thus "I-see to-John" for 'I see John' [Slavic*, Romance]

f. double determination in deixis (= a demonstrative adjective with a definite article and a noun (i.e., "this-the-man" for 'this man')) [Alb*, Grk, Slavic*]

g. possessive use of dative (genitive) enclitic pronouns (e.g. "brother to-me" for 'my brother') [Grk, Slavic, Romance]

h. the use of verbal forms to distinguish actions on the basis of real or presumed information-source, commonly referred to as marking a witnessed/reported distinction but also including nuances of surprise (admirative) and doubt (dubitative), e.g. "you-are (ADM) Albanian?!?!" [Alb, Slavic, Trk, Aromanian*]

i. the reduction in use of a nonfinite verbal complement ("infinitive") and its replacement by fully finite complement clauses, thus "I-want that I-go" (vs. standard European "I-want to-go") [Alb*, Grk, Slavic, Romance, Romany]

j. the pleonastic use of weak object pronominal forms together with full noun phrase direct or indirect objects ("object doubling", e.g. "him I-see John") [Alb, Grk, Slavic, Romance, Romany]

k. the formation of the "teen" numerals as DIGIT-'on'-TEN [Alb, Grk*, Slavic, Romance]

l. the presence of a (stressed) mid-to-high central (thus, schwa-like) vowel [Alb, Slavic*, Romance]

m. the presence of *i-e-a-o-u* in the vowel inventory without phonological contrasts in quantity, openness, or nasalization [Alb*, Grk, Slavic, Romance, Romany]

n. raising of mid-vowels ([+mid] > [+high]) [Alb*, Grk*, Slavic, Judezmo]

o. voicing of voiceless stops after nasals (NT > ND) [Alb, Grk, Aromanian]

p. presence of *ð θ (γ)* [Alb, Grk, Slavic*, Aromanian]

q. elimination of palatal affricates in favor of dentals [Alb*, Grk, Aromanian, Romany*]

Another view of the same data—many features represent divergences from earlier stages of the relevant languages, i.e.:

- **Modern Greek diverges from Ancient Greek, e.g. regarding future formation and use of infinitives**

- **Balkan Romance diverges from Latin, e.g. regarding use of weak (enclitic) pronouns**

- **Balkan Slavic diverges from Old Church Slavonic, e.g. regarding noun cases and admirative verbs**

- **Romany diverges from Old Indic (as represented by Sanskrit), e.g. regarding future formation**

- **Albanian diverges from reconstructed Proto-Indo-European, e.g. regarding definite article use (*no* article in PIE)**

And, there is enormous influence on the lexical level, mostly involving loanwords that are <mark>consistent with what is known about contact in the Balkans</mark>, the contact that gave rise to the structural convergence (and thus revealing the sprachbund)

- **Hypothesis: Sprachbunds arise under conditions of sustained, intense, intimate contact among speakers, with mutual multilateral multi-directional multi-lingualism**

A key distinction in Balkan language contact (as shown through the lexicon):

• object-oriented interaction (with a particular goal in mind, such as obtaining something (such as information), engaging in a commercial transaction (buying and selling), etc.)

• human-oriented interaction (no particular goal beyond the interaction itself, involving (generally) amicable interaction on a personal level)

Some standard typologies of loanwords

i) focusing primarily on the FORM of the loan (cf. Winford 2003):

• Haugen (1950): importation vs. substitution ("based on the presence or absence of foreignness markers")

• Haugen (1953): lexical borrowings ("imitation of some aspect of the donor model) vs. creations ("entirely native [with] no counterpart in the donor language" (but based on some nonnative material)

Cf. also Betz 1949: loanword vs. loan coinage

ii) focusing primarily on the CONTENT of the loan:

- **Bloomfield 1933 – cultural borrowings (those arising via the exchange, often mutual, of terminology between speakers of different languages, representing different cultures) vs. intimate borrowings (those not obviously linked to cultural objects)**

iii) focusing primarily on the MOTIVATION for the loan:

- **Hockett 1958 -- need borrowings (essentially Bloomfield's cultural type, though the motivation of "needing" a word for a (new) cultural item is at issue) vs. prestige borrowings (where the motivation is the "prestige" that the borrowing language speakers accord to material from the donor language)**

LATIN ==> Balkan languages (BS = Balkan Slavic / BR = Balkan Romance)

Lat *altar* 'refuge, protection': BS *oltar* 'altar' (OCS *olъtarь*), Alb *altar* (a later loan, cf. earlier layer borrowing *ltar*)

Lat *acetum* 'vinegar': BS *ocet* (OCS *осьтъ*), Alb *uthull*, Romn *oţet* (< BS)

Lat *camisia* 'shirt': MdvlGrk *kamisi*, MGk *poukamiso* 'shirt', Alb *këmishë* 'shirt', Rmn *cămaşă* 'shirt',

Lat *fossatum* 'military trench': MGk *fousato* 'army' Alb *fshat* 'village' (originally "fortified settlement"), Rmn *sat* 'village' (ditto)

Lat *furca* 'fork': Rmn *furcă* 'pitchfork', BS *furka* 'spindle, Alb *furkë* 'pitchfork, spindle', MGk *fourka* 'gallows'

VLat *furnu* 'oven': MGrk *fournos* Arom *furnu*, BS *furna*, Alb *furrë*, Tk *furun*

Lat *hospitium*: MGrk *spiti* 'house', Alb *shtëpi* 'house'

Lat *pomum* 'apple': Alb *pemë* 'tree', Rmn *pom* 'tree'

Lat *sagitta* 'arrow': MGrk *saïta* 'dart', Alb *shëgjetë* 'arrow', Rmn *săgeată* 'arrow'

Lat *scutum* 'shield': MGrk *skouti* 'thick woolen fabric', Alb *shqyt* 'shield', Rmn *scut* 'shield'; OCS *štitъ* 'shield' (= BS *štit*) is cognate.

GREEK

Grk *agiasma* 'santification':, Blg *agiazma/ajazma*, Mac *ajazma* 'holy water' Alb *ajazmë*, BR *aghiazmă* 'holy water'

Grk *anaphora* 'blessed bread': OCS *(a)nafora*, BS *nafora* 'holy or toasted bread' Alb *naforë*, BR *(a)naforă*

GK *anathema* 'curse, excommunication' BS *anatema* (also Mac *natema go* 'damn him', etc.), BR *anatemă*

Grk *eikona* 'icon': OCS *ikona*, BS *ikona*, Alb *ikonë*, BR *icoană*

Grk *kalogeros* 'monk': OCS *kalogerъ*, Blg *kaluger*, Alb *kallogjër*, BR *călugăr*

Grk *ēgoumenos* 'abbot': OCS *igumenъ*, Blg *igumen*, Mac *egumen*, Alb *(i)gumen*, BR *egumen* (*igumen*)

ROMANCE (French during Crusader era, in Medieval times, and others later, into early modern era):

French *exemple*:   (Middle) Greek *eksomplion* 'example'
French *monsieur*:   (Middle) Greek *misir* 'sir'
French *chambre*:   (Middle) Greek *tsampra* 'room'
French *roi*:   (Middle) Greek *roe/roi* 'king'

Venetian *veludo*:   Greek *veluðo* 'velvet' (vs. Itl *velluto*)
Venetian *zogo*:   Greek *tzoγos* 'gambling' (vs. Itl *giuoco*)
Venetian *aidar*:   Greek *aiðaro* 'help' (vs. Itl *aiutare*)
Venetian *cusina*:   Greek *kouzina* 'kitchen' (vs. Itl *cucina*)

Itl/Vntn *ruga*:   Albanian *rrugë* 'road'
Venitian *canal*:   Albanian *kanal* 'canal'
Italian *cassella*:   Albanian *kasellë* 'storage chest'
Italian *frate*:   Albanian *frate* 'monastic brother, friar'

TURKISH

aga '[Trk] lord' (MStTk *ağa*): Grk *agás*, Alb *aga*, BS/BR *aga*

asker '[Trk] soldier': Grk *askéri*, Rmni *askeri*, Rmn *ascher* (arch.)

cami 'mosque': Alb *xhami*, Grk *dzami*, Mac *džamija*, Arom *ğimie*

çorba 'soup': BS, Romani, Jud. *čorba*, BR *ciorbă*, Gk. *tsorbas*

çorap '(wool) stocking': Alb *çorap*, Grk *tsourápi*, BS, Jud. *čorap*,

      BR *ciorap*

dukkân 'shop': Mac *dukjan*, Blg *djukjan*, Alb *dyqan*

hendek 'ditch': Blg *hendek*, Mac *endek*, Grk *xandaki*, Arom *endec/hãndac*, Romn *hinidchi/hendechi*, Jud *jendek*

imam '(Muslim) priest': Alb *imam*, Grk *imamis*, Mac/Arom *imam*

minare 'minaret': Alb *minaret*, Grk *minares*, Mac *minaret*, Arom *minaree*

sokak 'street': Alb *sokak*, BS *sokak*, BR *socac*, Rmi *sokako*, Grk *sokáki* ('alley' in some)

tencere 'cooking pot': Alb *tenxhere*, BS *tendžere*, Grk *tentzeres*, Arom *tengire*, Romn *tingire*, Jud. *tenğere*

Some failings with these typologies:

a. by focusing on form, Haugen's does not build in the social context for the loans

b. the types listed above are not necessarily discrete – a cultural/need loan might be undertaken for reasons of (Hockettian) prestige or be associated with (Bloomfieldian) intimate contact

c. noncultural/nonneed loans are not always a matter of prestige, at least not obviously so; e.g.: Albanian (Arvanititka) diminutive –*zə* in Megarian Greek *liɣaz̲a̲* 'a little' [or covert prestige, e.g. Romani]

30

d. moreover, borrowing implies interaction between/among speakers, but … "non-interactive" borrowing does occur, e.g.:

- learned borrowings, as with Latin into Romance, or Church Slavonic into Bulgarian and Russian, or even Modern Greek *katharevousa* (high-style, Ancient-Greek-based) words adopted into *dimotiki* (low-style, colloquial Modern Greek)

- constructed neologisms (as with lots of western medical and other older technical terminology, generally with Greek or Latin roots, e.g. *encephalography* 'imaging of the brain', *electroencephalography* 'electronic encephalography', and *electroencephalographologist* 'a specialist who studies electroencephalographs'

What these typologies are missing:

- the full dynamics of the environment in which the borrowing occurs especially as to the MEDIUM through which the borrowing takes place

- this is a particular concern for the Balkans, since the lexical side of the Balkan sprachbund is only one dimension to the contact-related effects, inasmuch as there is massive structural convergence evident too (in the form of the future tense, in the use of finite subordination instead of infinitives, in the merger of genitive and dative case, etc.)

Motivating a New Type of Loanword (based on Friedman and Joseph forthcoming):

• Recognize a type of loan phenomenon which is consistent with what is known about contact in the Balkans, the contact that gave rise to the structural convergence (and thus revealing the sprachbund)

- sprachbunds arise under conditions of sustained, intense, intimate contact among speakers, with multilateral multi-/bi-lingualism which involves

- intimacy

- prestige

- multiple cultures (and thus "need" in some sense)

but more than that

35

Therefore, we need: a loan type that …

- **corresponds to these conditions**

and

- **is consistent with sprachbund formation**

Our suggestion:

- base loan typology on the mutual interaction, specifically on conversational interaction, between/among speakers

Therefore, we need to recognize a type of loans that are:

<u>E</u>ssentially <u>R</u>ooted <u>I</u>n <u>C</u>onversation

i.e.:

E.R.I.C. loans

These are loans that depend crucially on **speaker-to-speaker interaction** of an **on-going** and **sustained** kind, the sort of contact that can be characterized as **intense** and at the same time **intimate**, as opposed to **occasional** and **casual**.

Motivating the acronymic notion, formally and conceptually:

- formally, the acronym pays homage to Eric P. Hamp, the dean of Balkan linguistics, a long-time champion of the study of language contact in general and in the Balkans in particular

- conceptually, the notion offers a way to distinguish between loans that take place under sprachbund-conducive conditions and those that take place under casual contact situations.

To elaborate:

- face-to-face interaction, of the sort that would necessarily have occurred under the intense and on-going contact among speakers in the Balkans, is essential for creating and propagating the structural convergences typically taken as diagnostic of a sprachbund

- **and, that certain kinds of loanwords occur in such a social milieu is a bonus of sorts, and means that the loan words can be both an *indicator* that contact conducive to the formation of a sprachbund is taking place and a *by-product* of such contact.**

- **that is, these are loans that tell us something about speaker contact and about the sociolinguistics and the socio-history of the region**

- **this view draws on the notion of "degrees of contact", as recognized explicitly in the Thomason & Kaufman 1988 "scale of borrowability", where the borrowing of different types of linguistic material is claimed to correlate with different levels of intensity of contact among speakers.**

- **our contention is that, consistent with this scale, certain types of loanwords, especially those embedded in discourse and in conversational use and those that go beyond simple exchange of information, correlate with the intense, sustained, and intimate contact that is necessary for the formation of a linguistic area with structural convergence, i.e. a sprachbund.**

- these loans are "**sprachbund-consistent**", since they represent those lexical elements that most directly reflect the sort of language contact that is consistent with the emergence of a sprachbund, namely sustained intense contact on a day-to-day basis in a multilingual milieu.

- many ERIC loans are members of **closed lexical classes**, representing vocabulary domains that are generally held to be somewhat resistant to borrowing, and yet they are borrowed

Classes of ERIC loans (with examples to follow from all around the Balkans):

• **Kinship terms**

• **Numerals**

• **Pronouns**

• **Adpositions**

• **Negatives**

• **Complementizers**

• **Discourse elements (interrogatives, connectives, attitudinal expressives, interjections, gestures)**

- **Vocatives**

- **Onomatopoeia**

- **Reduplication (especially of an expressive nature)**

- **Expressive phonology**

- **Taboo & Abusive expressions**

- **Idioms (and phraseology more generally, incl. shared proverbs)**

- **"Intimate cultural items" (e.g. use of Muslim terms in 19th century Macedonia for Christian concepts, see below)**

- **Secret languages, trade languages, and jargonistic usage**

Examples of Loans in such Classes (a sampling, not exhaustive)

• **Kinship terms**

Turkish *baba* 'father' ➔ Albanian, Aromanian, Greek *baba* 'father' (also Pomak, archaic in standard Macedonian and Bulgarian)

Greek *patera* 'father' ➔ Aromanian (in Greece) *patera*

Turkish *nene* 'mother' ➔ Albanian *nënë*, Greek *nené*

Turkish *dayı* 'maternal uncle' ➔ Alb *dajë*, Mac *daja*

Greek *tsatsa* 'aunt' ➔ Aromanian *ţáţă* (surely related to Turk. *çaça* 'woman who keeps a brothel' but it may be from Grk)

- **Numerals**

Romani *efta, ohto, enna/enja* '7, 8, 9' (from Greek)

'11' – '19' as 'DIGIT-on-TEN', e.g. Albanian *pesë-mbë-dhjetë* '15' (widely cited in the literature as Slavic in origin but Albanian and Romanian disagree with Slavic in the gender of 'ten' in various formations, so this pattern may have a different origin and may in fact reflect some very early Albanian ("Albanoid") input to BR & pre-Alb/Slavic NWIE contact (cf. Hamp 1992)

Turkish numerals in Balkan languages (e.g. in Pomak (Muslim Bulgarian) in Greece today, digits for 'five' and above are all Turkish) also Christian Bulgarian from Thrace.

- **Pronouns**

Turkish *hiç* ➔ Macedonian *ič*, Albanian, Judezmo *hiç* 'nothing'

Greek *mu* 'my' ➔ Aromanian *–m* (vs. native *–n'i* (from Latin *mihi*, presumably via *mnihi)

Turkish *bu* 'this', *kim* 'who?' ➔ Ottoman-era Adrianople/Edirne Greek (OtEdGk) *bu* 'this', *kim* 'who' (only in *bu kim* 'who (is) this?') [Ronzevalle 1911]

cf. Macedonian (Cepenkov) *Stambol bu*.

- **Adpositions**

Turkish *karşı* (*karşi* in Balkan Turkish) 'opposite' ➔Albanian *karshi* (also in Aromanian, BS, Ottoman-era Edirne Greek (in Arom & Gk as a postposition)

Turkish *gibi* 'like' (postposition) ➔ Ottoman-era Edirne Greek *gibi* 'like' (postposition)

possibly Greek *me* 'with', if a borrowing from Albanian *me*, since derivation of Greek from Ancient Greek *metá* is problematic (requiring several ad hoc assumptions)

- **Negatives**

Greek *mi* 'prohibitive negator' ➔ Aromanian *mi*

Greek *oxi* 'no' ➔ Southern Aromanian *ohi* (Vrabie 2000)

Trk *hiç* (<P) 'mere trifle; nothing' ➔ OEGreek, Blg, Rmi *hič*,
     Alb *hiç*, Rmn *hici*, Judzm *hich,* Mac *ič,* Armn *ici* 'nothing,
     never, [not] at all, ...'

Trk (<P) *ne…ne* 'neither…nor' ➔ Agia Varvara Romani

Upward head nod for negativity (Greek, Romanian, some BS, and Turkish) — Greek is a posible source, given what is known about Ancient Greek gestures and the fact that the distribution especially in Italy coincides with geographic limits of Magna Graecia; but cf. extension eastward to India) — also ingressive dental click for negation

- **Complementizers**

Greek *(h)oti* 'that' ➔ Macedonian *oti* 'that'

Greek *(h)o, ti* 'for that reason' ➔ Macedonian/Bulgarian *oti* 'because'

Turkish (< Persian) *zira* (st), *zere* (dial) 'because' ➔ BS *zer, zerem, zare, zerja, zeri*

Turkish (< Arabic) *zahir* (lrnd), *zar* (colloq) 'apparently. evidently' ➔ BS *zar*, Armn *zări,*

- **Discourse elements (interrogatives, connectives, attitudinal expressives, interjections, gestures)**

NB: These are ERIC forms, since they serve as the "glue" of everyday interactions between people, and include markers of common discourse strategies (e.g. questioning) as well as frequent discourse "particles" and indicators of an individual's status relative to other interlocutors (e.g. solidarity, distancing, etc.) but also those that serve a purely expressive purpose (i.e., elements that add "color" to conversation):

INTERROGATION

Trk *mi/mü/mı/mu* '(focalizing) marker for yes-no questions' ➔ Ottoman-era Adrianople Greek *mu* 'marker for yes-no questions', Romani *mi* & *li* as interrogatives and dubitatives

-- single harmonic form borrowed (adopted without adaptation)

-- postpositive in OtEdGk & Rmi, as in Trk, though slightly different (phrase-final in OtEdGk, post-focus in Trk & Rmi)

-- has other, more discourse-based functions, in OtEdGk & some Rmi: irony, doubt, challenging, swearing (an oath)

--examples:

boreís mou
can/2sg   Q
'Can you (do it)?'
(Standard Greek: *μπορείς;*)

tha 'rt's   mou
FUT come/2sg Q
'Will you come?'
(Standard Greek: *ta [e]rtheis*)

Tu li/mi sijan ki Amerika! Hohavea! [KumanovoAril/Barutči]
you Q are in America.   You.lie

O Devla,   ta   oda mandar da   butrašadi   isja-li [Sliven]
O   God.VOC   and   they   me.ABL   and   more.scared
was/were.IMP-*li*
  'Oh my God, they're more timid than I am!'

57

provincial Turkish (nonharmonic) *de* **'c'mon! now then' & BS**
  de **(cf. Russ.** *de,* **Ukr** *di***, etc) 'reported>emphatic'**
  (<dě[ja]ti)➡ Greek *de* **signaling impatience, Alb, BR** *de*
  emphatic with imperatives, etc.

Slavic *da* **'yes' ➡ BR** *da*

Albanian *po* **'yes' ➡ some West Aromanian** *po*

Greek *malista* **'yes (indeed)' ➡ East Aromanian** *malista*

CONNECTIVES

Mac (=Sl) *i* 'and' ➔ Aromanian, Romani

Trk *[h]em* 'and, too, and yet' ➔ OtEdGk, Alb, BS, Armn, Romani

Trk *amma* ➔ *ama* Alb, **BR**, BS, **Gk**, Rmi, Jdzmo

Trk *ya … ya* 'either … or' ➔ Alb, **Armn**, BS, **Gk**, Romani

Mac *ili* (= Sl) 'or' ➔ WRT (Macedonia)

Mac (= SSl) *pa* 'and so, well' *(< pak = ChSl paky)* ➔ Alb

Trk acaba 'I wonder if; oh indeed!'

Blg *adžaba/adžeba* 'I wonder; is it so?', **Mac** *adžaba*, **Gk** (of Edirne) *adžiba*, **Armn** *hagiba/ageaba* 'is it so?', **Megl** *adžaba*

Gk aliθeia 'truly? really?'

Armn *aliθea/alihea*

Trk bari(m) 'at least; for once'

Alb *bar/bare(m)/bari*, **Armn** *báre/bári/bárim*, **BS** *bar/bare/bárem/barém/bári/barí/bárim/ barím*, **Rmn** *barem*, **Gk** (dialectal) *barim*, **Rmi** *barem*

Trk belki(m) 'perhaps, maybe'

Alb *belqim*, **Armn** *belchi*, **BS, Jdzm** *belki* 'maybe; probably; as if', **Gk** *belki(m)* (dialectal)

Trk değil mi 'isn't it so?'   Alb *dilmi* 'since; because; after', Armn *delme* 'since', BS

         & Rmi *delmi/dilmi/dilma* 'isn't it?'

Trk demek 'it means,   BS, Alb, Rmi *demek* 'it means; supposedly (indicating

   as if'   disbelief)'; also N. Gk

Trk elbet(te) 'certainly,   Alb *(h)elbet(e),* **Armn** *elbet(e)* 'possibly; assuredly', BS

   surely'   *(h)elbete/elbetta/helbette/helbet(t)ja*, Rmn *(h)elbet*

Trk gûya   'as if;   Alb *gjoja/gjyja,* **Armn** *ghio(i)a/ghio(i)am/ghioae*, BS

   supposedly'   *g'óa/g'oé/gjúva/g'oj/g'óve*, **OtEdGk** *γ'a*

Alb mbase 'perhaps;   **Gk** *(m)bas (ke)* 'perhaps'

   maybe'

Trk sanki(m) 'as if'   Armn *sanki,* BS *sankim* 'actually; that is to say; as if',

                   Rmn *sanche/i,* **OtEdGk** *sangim*

OTHER SENTENCE ADVERBS / MITIGATING MODIFIERS

Trk bile 'even; already'

Alb *bile* 'even; in fact', **Armn** *bile*, Blg *biljá(m)/bilé(m)*, Rmi *bila(m)/bilim*

Trk sade 'only, simply'

Alb *sade*, **Armn** *sade*, BS *sa(a)dé*, Rmn, OtEdGk *sadé*, Rmi *sáde/sadé*

Trk salt 'only'

Alb *sall(a)/sallde/sallte* (dialectal), BS *sal/sált(e)*, MRmn *sal/săl*, Rmi *salt/salde*

Trk tamam 'just right; there you have it!'

Alb *tamam/taman*, **Armn** *tamam/tamamá tamamaná*, BS *tamám/tamán*, Rmn (dialectal) *taman*, **Gk** (dialectal) *tamam(i)*, Megl *tamam*, Rmi *tamami*

Rmn mai 'almost'

Blg *mai*

'unceremonious term of address' (lots of variants, almost all ultimately from Greek, where dialectally some 55 different forms are found, almost all stemming from AncGk *more* (vocative of 'foolish'; cf. Joseph 1997):

Turkish:	*bre, bire, be*
Albanian:	*ore, or, mor, more, moj, ori, mori, moré, mre, voré, bre, be*
Romanian:	*bre, mă, măi*
Balkan Slavic:	*more, mori, bre, be, or...*
Romani:	*more, mori, bre, be, or...*
Aromanian:	*more, mori, bre, be, or...*
Meglenoromn:	*more, mori, bre, be, or...*
Judezmo:	*bre*
Greek:	*moré, bre, vre, re, aré, maré, marí, oré, voré, etc.*

EXHORTATIVES

Turkish *(h)ay de* ➔ **Alb, BR, BS, Rmi, Jud** *[h]ajde*, **Greek** *aide*
   'c'mon'

Greek *ela* **'c'mon'** ➔ **BS, Alb** *ela*

Trk *ha* **'a call to action'** ➔ **OtEdGk** *ha*

GREETINGS

Trk *merhaba* 'hello'➡ Blg *maraba/merhaba*, OtEdGk;
   Alb&Mac in Gora *merhaba*

Gk *geia sou* ([ja su]) 'hello' (lit.: "health to-you") ➡ **Armn
   γa su** = BS *zdravo, zdravi bili* (cf. Russian *zdravstvujte*) cf.
   Alb *tungjatjeta,* Armn *s-hits ghini*

--(directionality unclear):
   Greek *ti kaneis? kala* 'what do-you? [I-am] well (ADV)
   = Albanian *ç'ka po bën? mirë.* (*mirë* as adverb)
   = Macedonian *š[t]o pra[v]iš? dobro.* (adverbial form)
   = Bulgarian *kakvo praviš? dobre.* (adverbial form)
   = Romani *so kere[s]? shukar.*
   = Romanian *ce mai faci? bine.*
   = Aromanian *tsi fats? gine.*
   = Turkish *n[e] [y]aparsın? iyi.*

- **Vocatives**

Blg preservation of VOC until late 20th century

Albanian *biro* 'O son!', *Agimo* 'O Agim!", with *–o* apparently from Slavic vocative

Romanian vocative endings influenced by Slavic, cf. especially *–u-le* (PL), apparently with Slavic vocative particle *–le*

Trk *oğlu-m* 'son-my' (term of endearment) ➔ [19th cent.] OtEdGk *oɣlum*, Rmi *olum*

Trk *can-ım* 'soul-my" (term of endearment) ➔ OtEdGk *džanım*, Pontic *dźanum*, Blg *džanăm*, Mac *džanam*, BCSM *džanum*, Alb *xhanëm*, Armn *geanãm*

- **Onomatopoeia** (*vel sim.*)

'dog noise':   Albanian *ham-ham*, Romanian *ham*, Greek *γav γav* , Turkish *hav hav*, Macedonian *av av*, Romani *hau-hau*, Bulgarian *bau-bau*

'knocking': Albanian & BS *tak-tak*, Greek *tak tak*

'noise for attracting a cat': Greek *psi psi psi*; BS, BR *ps ps ps*

- **Reduplication** (especially of an expressive nature)

Turkish affective *m*-reduplication (e.g. *kitap-mitap* 'books and such')

Alb   *çikla-mikla* 'tiny bits and pieces; crumbs; trivia'
BS   *knigi-migi* 'books and such'
            kal-mal 'mud or whatever'
Rmn   *ciri-miri* 'confusion'
Jud   *livro mivro* 'books and such'
            sapatos mapatos 'shoes, shmoes' (Bunis 1999)
Gk   *tzantzala mantzala* 'rags and such'
            ta sandala ke ta mandala 'stuff and things' (OtEdGk)
            surðu murðu 'topsy-turvy' (Levkas dialect)
            ara mara 'who cares?'
            ares mares (kukunares) 'nonsense'

Turkish-style CVC- intensive prefixing (e.g. *beyaz* 'white' /
**   *bem-beyaz* 'pure white'):**

Greek *tsiplakis* 'naked' / *tsir-tsiplakis* 'stark naked'

Bulgarian *baška* 'other / *bambaška* 'very particular'
    gol-goleničăk 'stark naked'

Macedonian *gol-goleničok* 'stark naked'

Romanian *gol-golut* 'stark naked'

- **Idioms** (and **phraseology** more generally, incl. shared proverbs)

Friedman 1986 on Macedonian:

> "the use of *jade* 'eat' to mean 'be subjected to' as in *jade k´otek* 'get a beating' (literally 'eat a blow') is obviously based on Turkish *kötek yemek* 'ibid.', and so *jade dožd* 'get soaked' ('eat rain', Tk *yağmur yemek*) or *jade gomno* 'say something stupid/embarrassing' ('eat excrement', Tk *bok yemek*, cf. English *to put one's foot in it*) can likewise be identified as calques."

- **"Intimate cultural items"**

-- **Muslim terms in 19th century BS for Christian concepts**

kurban (< Turkish *kurban* 'sacrifice') 'eucharist'

Sajbija (< Turkish *sahib* 'master') 'the Lord'

Kurtulija (< *kurtul-* 'save') 'the Saviour', *kurtulisvam* (< Tk. *kurtulmak*) 'to save'

hadžija (< Turkish *hacı*) 'Christian pilgrim to Jerusalem' (and not only Muslim pilgrim to Mecca)

- **Secret languages, trade languages, and jargonistic usage**

Capidan 1940: "special language" of tailors in a Greek village in the Pindus who otherwise speak Greek but, in this jargon, use a language variety that is largely Aromanian in vocabulary

Armn and Alb ➜ Mac & Gk sheepherding terminology

- **Taboo expressions / Curses   (NB: for the most part, these tend to be inherited, not borrowed but there is some borrowing):**

NOUNS - BODY PARTS
prick/dick/cock   c*nt
balls   t*ts
ass/arse

VERBS - ACTIVITIES
f*ck   suck   (irrumate)   jack-off

NOUN&VERB - BODY PRODUCTS
shit   fart   piss   cum

PEOPLE
bastard - bitch
faggot - dyke
whore/[hustler] - pimp

73

Some examples of such borrowings:

mindzh 'c*nt' (Rmi > Mac, but not Blg)
pizda (Slv; BR > *kizda*)
kar? (Rmi & Alb)
chorajbe 'cum' (Rmi > Mac)
kurva 'whore (Slavic > all East Europe from Estonia to Greece)
bulangiu / bulash / bythexhi 'faggot' (Rmn < Rmi = Alb)

Tk puşt = Alb pusht = Gk πούστις ≠ Rmn puşt ⊐ BS, Armn, Rmi [Blg&Mac 3(1)-pušarak)

BS peder (Mac > per) 'f*ggot' > 'bastard'

pezevenk -Mac-3,4; Blg-3(1), πεζεβεγκης (1-Cyprus), Rmn-4; Alb pizeveng-1, Armn-4 'pimp' > 'bastard'
orospu - Alb & Mac [orospija] -1; Blg-3(1), Rmi-orospu-2, Rmn-4; Armn-4 Gk-3,4 (NB Gk *putana*) 'whore'
sefi - Mac-3,4 Rmi-4 Gk-4, Rmn-4 Blg-4, Armn-4, Alb-4 'lech'
sikter Mac-1,2, Blg-3(1), Rmn-1, ; Armn-1; Rmi 2, Gk σικτιρ 1,2; Alb 2 'get f*cked' > 'scram'

74

Turkish *siktir* **'get f*cked'**

Judezmo *asiktar senda*, verb *asikterear*

Greek *asiktir – asikhtir (more expressive); asiktirizo = trans'v;*
siktir pilaf

Macedonian *sikter* 'scram' cf. also *aneni sik'im < anneni*
sikeyim; siktir kafe

Trk *annana babana* 'to your father, to your mother' ➜
(OtEdGk) αννανά μπαμπανά (vulgar expression)

One final clearly conversational example:

• **an expression used by some Greeks (maybe obsolete now but it did exist) in the game of "peek-a-boo" that adults and older children play with young children:**

  buli buli buli buli … dza

• **the *buli* part (repeated four times) comes when the face is covered up by one's hands and the *dza* part comes when the hands open up to reveal the face.**

Where does this come from?

• in Greek, *buli* is just a nonsense word

• *dza* (with variant *tsa*)= an interjection marking surprise or
indicating something like "here I am (somewhat
unexpectedly)"

But where does that *dza* come from?

• the only dictionary to comment on its source says it is a "nursery word", and its use and form do make sense in that regard, based on special expressive status claimed for Greek [ts]/[dz] (Joseph 1982, 1984, 1994 and elsewhere)

   --these sounds are said to be "allolinguistic", i.e. showing a marked functional status for linguistic elements involving being on the margins of "core" information-oriented communication (Wescott 1976)

- still, even with an allolinguistic rationale for *tsa*/*dza*, one has to ask why, if the word has a nursery-related origin, it has the particular form that it does.

- but, there is a compelling source for *tsa*/*dza* from a language outside of Greek:

- Albanian has an interjectional word spelled < xa > (phonetically [dza]) that, as listed in Mann 1948, has a meaning 'here you are'.

--this presentational meaning fits in well in the game of peek-a-boo as the sound that accompanies the revealing of the face

--so, if this source of the Greek utterance is accepted, this form would have entered Greek either through direct contact with Albanians, e.g. in northwestern Greece, or through Arvanitika, the Albanian dialects spoken mainly in Central Greece, the Peloponnesos, and Attica.

But what about the rest, the *buli* part?

• **that too has a compelling source in Albanian:**

     -- [buli]: from Alb verb *mbyll* 'close, shut' (3SG.PAST)

• **thus, the phrase in its etymological meaning would be:**

     -- "[when the hands cover the face] (it-has-)closed,
        (it-has-)closed, (it-has-)closed, (it-has-)closed …
        [when the hands open up to reveal the face] Here-it-is!".

• **Phonetically the Greek [buli] would be an expected rendering of a borrowed Albanian *mbyll*:**

  -- the initial *b*- for Albanian [mb] conforms to prevailing Greek phonotactics at the likely time of borrowing (with voiced stops without a nasal "prop" being allowed in word-initial position)

  -- the [u] for the Albanian front rounded [y] vowel accords with Greek lacking an [y] at this time (NB: even if Arvanitika is the source, with [i] for [y] elsewhere in Alb, since Grk has [u] for earlier [i] in labial contexts).

- **This expression would have been borrowed only through use, presumably by Arvanitika speakers using it to play with Greek babies (and thus made available to Greek speakers) or by Greeks over-hearing Arvanitika speakers using it with their own babies.**

Either way, the conversational element and 'intimate' contact are paramount here.

By way of conclusion ....

• these facts shed light on the particulars of the language contact situation in the Balkans and give us a better basis for understanding the sprachbund, while at the same time motivating a new class of loanword

• moreover, they show which parts of the lexicon, even under conditions of extreme and intense contact, resist up to the bitter end (in particular, curses and obscenities) so that there is structural isomorphism in the partitioning of the lexicon in languages in contact