Small-scale patterns in a larger picture: The loss of feature specifications in the Balkan Slavic pronominal system

Sonnenhauser, Barbara ; Escher, Anastasia

DOI: https://doi.org/10.1515/if-2021-005

Indogermanische Forschungen

Zeitschrift für Indogermanistik und historische Sprachwissenschaft

Begründet von
Karl Brugmann und Wilhelm Streitberg

Herausgegeben von
Benjamin Fortson, Götz Keydana, Melanie Malzahn und Paul Widmer

126. Band
2021

DE GRUYTER

Alle für die Indogermanischen Forschungen bestimmten Beiträge reichen Sie bitte über das Einreichungsportal ein unter: https://mc.manuscriptcentral.com/indo

oder senden Sie per Briefpost an:
Sprachwissenschaftliches Seminar der Georg-August-Universität Göttingen,
Indogermanische Forschungen, Käte-Hamburger-Weg 3, D-37073 Göttingen.

ISSN 0019-7262
e-ISSN 1613-0405

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de abrufbar.

Druck und Bindung: CPI books GmbH, Leck

www.degruyter.com

Inhalt

Barbara Sonnenhauser and Anastasia Escher

Small-scale patterns in a larger picture

The loss of feature specifications in the Balkan Slavic pronominal system

Abstract: Telling apart family-internal developments from contact-induced changes and socio-geographically conditioned areal trends is still one of the main challenges in tracing the development – and stability – of languages. Notably prima facie contact phenomena might turn out to have resulted from the interaction of a variety of sources upon closer inspection; assessing their interaction is a crucial requirement for comprehending the dynamics of change. Based on the example of simplification patterns in the system of short personal pronouns observed in the Balkan Slavic dialects located in the territory of North Macedonian and Greece, the present paper illustrates the necessity of identifying the multiplicity of triggers and their interaction, thereby making a case for the relevance of small-scale, transient patterns in understanding diachronic processes. In particular, it places the simplifications within the pronominal system into the larger context of differential object indexing by preverbal pronominal elements in the Balkan languages. With the loss of agreement distinctions on pronouns being characteristic for the last phase in the emergence of object indexing, both processes appear to be closely related. Obviously, in this case, the convergence of various different processes has created a favorable environment for the mutual reinforcement and stabilization of two otherwise highly volatile phenomena.

Keywords: 3rd person pronouns, agreement features, object indexing, variation, Macedonian, Balkan area

1 Introduction

The Balkan Slavic dialects located in the territory of North Macedonia and Greece display remarkable tendencies of simplification in their system of short personal pronouns: the form *mu* is used for all singular and plural referents expressing 'dative' relations in the southern varieties, and the form *ga* for singular referents expressing 'accusative' relations in all genders in the northern varieties.

Barbara Sonnenhauser, Slavic linguistics, U Zurich; barbara.sonnenhauser@uzh.ch
Anastasia Escher, Slavic linguistics, U Zurich; anastasia.makarova@uzh.ch

https://doi.org/10.1515/if-2021-005

In Standard Macedonian, the system of personal pronouns comprises long (stressed) and short (clitic[1]) forms. All of them distinguish number (SG and PL) and, for the 3SG, gender (M/N vs. F), see Table 1.

Table 1: Personal pronouns in Macedonian.

	NOM	DAT.long	DAT.short	ACC.long	ACC.short
1SG	*jas*	(*na*) *mene*	*mi*	*mene*	*me*
2SG	*ti*	(*na*) *tebe*	*ti*	*tebe*	*te*
3SG.M	*toj*	*nemu*/*na nego*	*mu*	*nego*	*go*
3SG.F	*taa*	*nejze*/*na nea*	*i*	*nea*	*ja*
3SG.N	*toa*	*nemu*/*na nego*	*mu*	*nego*	*go*
1PL	*nie*	*nam*/*na nas*	*ni*	*nas*	*ne*
2PL	*vie*	*vam*/*na vas*	*vi*	*vas*	*ve*
3PL	*tie*	*nim*/*na niv*	*im*	*niv*	*gi*

Just like the long forms, the short ones may be used referentially, i.e. refer to an already introduced or situationally given referent, see (1).

(1) a. *Im rekov.*
DAT.3PL say.AOR.1SG
'I told them.' makedonski.info

b. *Í pomognav.*
DAT.3SG.F help.AOR.1SG
'I helped her.' makedonski.info

c. *gi čitav i ...*
ACC.3PL read.AOR.1SG and
'I read them and ...' makedonski.info

In addition, the ACC and DAT short forms function as indices for direct and indirect objects (henceforth: IO and DO), occupying the immediate preverbal position, (2a), except for imperatives, (2b) and participles, where they follow the verb (Bužarovska 2001: 2). Prosodically attaching to the verb, regardless of its position within the clause, the short forms employed as object indexes are part of the verb phrase.

(2) a. *[Svojot život]$_i$ ḱe í$_j$ go$_i$ posvetam na [muzikata]$_j$*
my life FUT DAT.3SG.F ACC.3SG.F dedicate.PRS.1SG to music.SG.F.DEF
'I will dedicate my life to music' makedonski.info

1 The term "clitic" encompasses both prosodic and morphosyntactic aspects. In this paper, we deal only with the morphosyntactic.

b. *Dajte* *í*$_i$ *šansa na [mladinata]*$_i$
give.IMP DAT.3SG.F chance to youth
'Give youth a chance' makedonski.info

Dialectal data from the mid-20th c. published in Vidoeski 1999 display two patterns of simplification in the pronominal system, namely the generalization of specific forms resulting from the loss of feature distinctions. As argued in this paper, these patterns emerge from the interaction of various conditions that favor in sum the mutual reinforcement and temporary stabilization of two developments that are otherwise volatile in Indo-European: the loss of gender/number distinction for referentially used short personal pronouns and object indexes developing into purely syntactic markers without any referential feature specifications. The data discussed in this paper thus illustrate the need to consider the full complexity of interacting causes in tracing the processes of language change instead of resorting to only one cause, even if prima facie evidence and the pursuit of elegance in description might tempt one to do so.

The paper is organized as follows: the patterns of pronoun generalization are illustrated in section 2. Section 3 discusses possible sources for these changes, ultimately arguing for an interplay of factors. The relevance of these small-scale and short-lived patterns encountered in Macedonian from the larger perspective of diachronic processes in Indo-European is indicated in section 4; section 5 provides a short conclusion.

2 Patterns of pronoun generalization

Vidoeski's (1999) corpus is the only published collection of dialect texts covering the whole territory of Macedonia, illustrating every dialect type with at least a small sample. The texts were collected in the second half of the 20th c., i.e. shortly after the codification of Macedonian. Thus, informants had not yet been subject to the systematic implementation of the standard variety.

These data display two major patterns of short pronoun generalization, which are distributed across some of the main dialect groups (see Fig. 1).

1. Southern Aegean dialects: Generalization of *mu* as a marker for 'dative' relations, in particular indirect objects, for all genders in SG and PL.
2. Northern dialects: Generalization of *ga* (with *go* and *gu* as variants) as a marker for 'accusative relations', in particular direct objects, for all genders in SG.

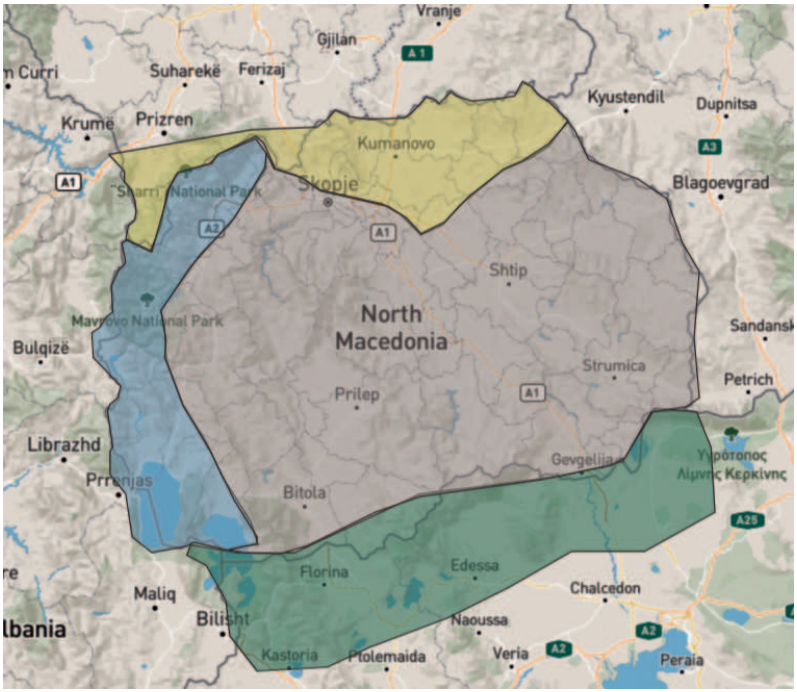


Fig. 1: Dialectal distribution of generalization patterns (blue area = instable generalization in the western dialects, yellow area = *ga* generalization in the nothern dialects, green area = *mu* generalization in the southern dialects)[2]

Peripheral western dialects are characterized by an inconsistent generalization of *mu* and *ga* forms.

The generalization of *mu* concerns 'dative' relations, i.e. the expression of indirect objects and internal possessors. While *mu* as the exponent of dative relations for masculine singular referents is used across the South Slavic continuum (including the standard languages), 'dative' relations for feminine referents feature various exponents. The Macedonian standard language employs the central west Macedonian form *i*, which is also the form closest to the proto-Slavic DAT.SG.F **jeji* (OCS *ѥи*; Schenker 1993: 90). In the eastern dialects, *in* and *vu* are encountered (for their origin see Seliščev 1918: 197; Belić 1905: 440–441; Vidoeski 1965: 61), in the West, *je* (*e*) is competing with *mu*, which in turn prevails in the South, see Fig. 2 (p. 89).

2 For the visualization of the data, the following R packages were used: ggmap (Kahle & Wickham 2013), leaflet (Cheng, Karambelkar & Xie 2020).

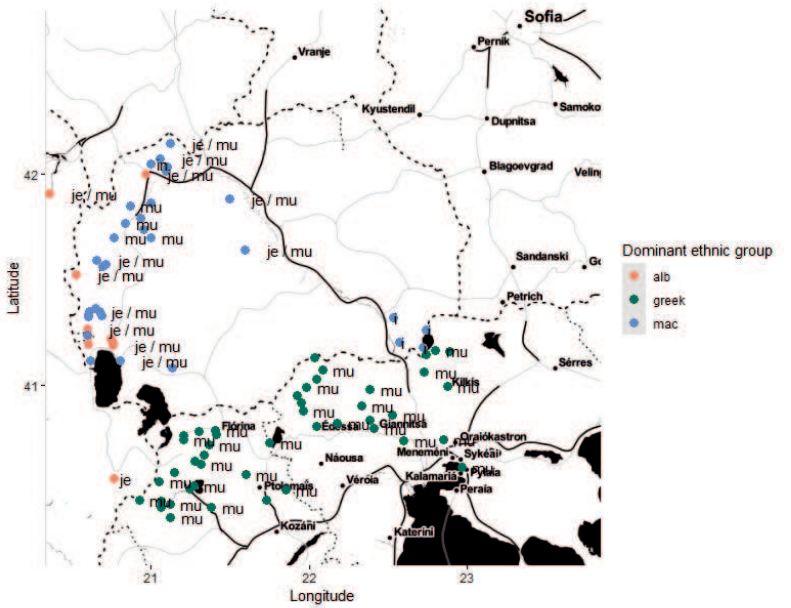


Fig. 2: Short pronominal forms indicating dative relations of feminine referents in western peripheral and southern Macedonian dialects

The southern Aegean Solun-Voden and Kostur-Korča[3] dialect groups (colored in green on the map in Fig. 1, p. 88), i.e. those dialects that exhibit a prevalence of *mu* for feminine referents, are located in Greece (except for the Korča dialect which is situated in Albania) and are thus isolated from the influence of the major Slavic-speaking area, 20th century Yugoslavian language policy and the Serbian and Macedonian standards. They still find themselves under the pressure of Greek language and identity policy that prohibits speaking a Slavic idiom. In these dialects, the pronoun *mu*, which is otherwise specified as dative masculine singular across the South Slavic dialects, has been generalized as an indirect object marker for all genders in the singular, see (3), and also for the plural, see (4).[4] As a consequence, the pronoun *mu* operates as the only exponent of 'dative' relations across gender and number.

3 The names of the dialect macrozones are given according to the classification by Vidoeski (1999).
4 The earliest examples of the feature can be found in folklore texts from the 19th century (Cepenkov 1892).

(3) *Se radva Marko oti dete ima.* ***Mu***$_i$ *veli* ***na samovila-ta***$_i$ *...*
REFL be glad.PRS.3SG Marko that child.SG.N have.PRS.3SG PRN.GEN.IO say.PRS.3SG to fairy.SG.F.-DEF
'Marco is happy that he has a child. He says to the fairy...'
Vidoeski 1999: 357; Gorno Neolani, Kostur-Korča dialect group

(4) *Uče-še ama [[na tatko mu] i [na majka mu]]*$_i$ *ne* ***mu***$_i$ *kažva-še*
learn-IMPF.3SG but on father.SG.M PRN.GEN.IO and on mother.SG.F PRN.GEN.IO not PRN.GEN.IO tell-IMPF.3SG
'He learned but didn't tell it to his father and mother'
Vidoeski 1999: 362; Bapčor, Kostur-Korča dialect group

The generalization of *ga* to express 'accusative' relations, in particular direct objects, can be observed in the Gora and Kratovo-Kumanovo dialect groups north to the line Tetovo – Skopska Crna Gora – Kumanovo – Kratovo (yellow on the map in Fig. 1). This area is transitional between Torlak[5] and Macedonian, displaying a considerable amount of multilingualism with Macedonian being predominant. In these northern varieties,[6] the form *ga* and its phonetic variants *go* or *gu* (Tetovo: *ga*,[7] Vratnica: *gu*/*go*,[8] Gora: *ga*,[9] Skopska Crna Gora: *gu*,[10] Kumanovo: *gu*,[11] Kratovo: *gu*[12]), has been generalized as a direct object marker for masculine and feminine substantives in the singular, see *ga* in (5).

(5) *Ona otišla kăj neko-a žen-a što razbirala od magjiite i* ***ga***$_i$ *prašue žen-a-ta*$_i$*...*
she go.PERF.SG.F to some-SG.F women.SG.F that know.PERF.SG.F from magic.PL.DEF and PRN.GEN.DO ask.PRS.3SG woman.-SG.F.-DEF.F
'She went to some woman that knew magic and asks the woman...'
Vidoeski 1999: 39; Siričino, Tetovo dialect group

The long belt of peripheral western dialects (Kumanovo – Debar – Struga – Ohrid, colored in blue on Fig. 1, p. 88) constitutes a mixed zone. In its southern part,

5 Torlak is a balkanized dialect of southeast Serbian.

6 Dialects of northern North Macedonia are studied rather poorly; the Vidoeski corpus, too, comprises only a few illustrative texts. The most recent scientific studies in this region are related to ethnolinguistics (Plotnikova 2018).

7 Čelopek, Siričino, Preljubište, Jedoarce, Tearce.

8 Vratnica, Stare Selo, Dikance.

9 Dikance.

10 Kučkovo, Mirkovce, Ljubance.

11 Kumanovo, Orašac, Rugjince, Stepance, Petralica, Odreno, Dobrovnica.

12 Kratovo, Železnica, Lesnovo, Zletovo, Dolni Stubol.

instances of *mu*-generalization are attested, in the northern part cases of *ga*/*go*/*gu*-generalization. However, both tendencies are inconsistent, and the full pronominal system, as given in Table 1 (p. 86), can be observed as well.

3 Underlying processes

At first sight, the patterns described in section 2 seem to be straightforward contact phenomena. First, similar phenomena can be found in the neighboring non-Slavic varieties to the North, West and South, in particular in Albanian and the Struga variety of Aromanian, which is in contact with Albanian and Macedonian; second, the eastern dialects, which are not in contact with these varieties, do not show any generalization tendencies. However, upon closer inspection, further possible causes are equally likely to have contributed to this development. In particular, feature-specific and language-internal factors as well as cross-linguistic trends seem to have conspired with contact-related factors.

3.1 Contact: loss of feature distinctions

In the Balkan linguistic context, joint features have been traditionally explained as resulting from a centuries-long language contact of the attested Balkan languages and the "tendency towards symmetry and syntactic leveling between [...] language systems" (Bužarovska 2001: 13) or as a consequence of substrate/adstrate/superstrate relations that have characterized the complex contact history since the settlement of the Slavs in the Balkans. Against this background, the process of pronoun generalization seems to constitute another outcome of this particular contact situation.

The most obvious reason for considering language contact as a crucial trigger in the process of pronoun generalization is provided by the neighboring non-Slavic languages which exhibit the same phenomenon. In Albanian, neither accusative nor dative short pronoun forms distinguish gender in 3SG and 3PL. The full forms distinguish masculine and feminine only in dative singular and accusative plural (Buchholz & Fiedler 1987: 281), see Table 2 (p. 92).

Table 2: Dative and accusative personal pronouns in Albanian

	SG				PL			
	M		F		M		F	
	full	clitic	full	clitic	full	clitic	full	clitic
DAT	*atij*	*i*	*asaj*	*i*	*atyre*	*u*	*atyre*	*u*
ACC	*atë*	*e*	*atë*	*e*	*ata*	*i*	*ato*	*i*

In addition, generalization across number can be observed. According to Newmark, Hubbard & Prifti (1982: 25), in colloquial Albanian, the DAT.SG *i* can be found as plural indirect objects markers instead of the standard DAT.PL *u*, see (6).[13]

(6) *pionirët* i_i *dhuruan miqve*$_i$ *buqeta me lule*
pioneer.PL.DEF 3SG.DAT presented friend.DAT.PL bouquet.PL with flower.PL
'the pioneers presented bouquets of flowers to their friends'
Newmark, Hubbard & Prifti 1982: 25

Concerning Aromanian, the descriptions of different varieties as well as general overviews of this highly variable language (Ianachieschi-Vlahu 2001; Narumov 2001) present a full system of short pronouns with distinct forms for both genders. The phonetic realization of these forms might vary considerably in different varieties, see Table 3.

Table 3: The pronominal system of Aromanian varieties

	SG				PL			
	M		F		M		F	
	full	**clitic**	**full**	**clitic**	**full**	**clitic**	**full**	**clitic**
DAT	*a lui*	*lji/ãlj/l*	*a ljei*	*lji/ilj/lj*	*a lor*	*lã/ãl*	*a lor*	*lã/l*
ACC	*el/nãs*	*lu/ãl*	*ea/nãsa*	*u/o*	*elji/nãshi*	*lji/ãlj/lj*	*eale/nãse*	*li/le*

Recent fieldwork in the Struga and Bitola districts, an Albanian-Macedonian-Aromanian convergence area with Albanian predominance and sporadic *mu*-generalization for dative relations in Macedonian, reveals cases of pronoun gener-

13 While Newmark, Hubbard & Prifti (1982) interpret this as a generalization across number, it cannot be excluded that we are dealing here with an expansion of the ACC into the DAT in the plural, as suggested by an anonymous reviewer. In any case, it is the form covering the most functions within the paradigm already that has been generalized even further.

alization as well. Here, the formally ACC.SG.F pronoun is used not only for feminine referents, as in (7), but also for masculine ones, see (8) and (9).

(7) *Dad-a* ***u=***$_i$ *mvãscu* *[pi hiljãs-ã]*$_i$
mother-DEF.F ACC.3SG.**F** dress.AOR.3SG on daughter.F-INDEF.F
'The mother dressed the daughter' Arom.

(8) *Fet-a* ***u=***$_i$ *cunushtie* *[pi ficior-u]*$_i$[14]
girl-DEF.F ACC.3SG**(F)** know.IMPF.3SG on[15] boy-DEF.**M**
'The girl knew the boy' Arom.

(9) ***u=*** *vădzui* *asera* *tu ubor-u*
ACC.3SG.**(F)** see.AOR.1SG yesterday in yard-DEF.M
'I saw him yesterday in the yard' Arom.

These observations suggest contact with Albanian to be a main trigger of the gender-generalization patterns.

Insight into the potential role of language contact for the generalization of *ga* as a direct object marker can be gained from the ethnolinguistic situation in the regions in question, i.e. the northern dialect zone. Fig. 3 (p. 94) displays the information provided by the 2002 population census of Macedonia (Census) concerning the prevalent ethnic group for the relevant locations as covered by the Vidoeski corpus.

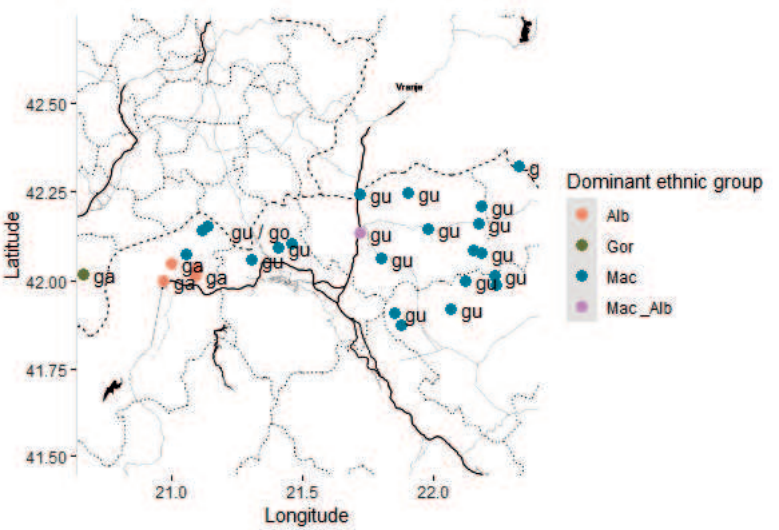
The data show that direct Albanian-Slavic contact is likely only in a limited number of locations. Albanians are predominant merely in the district of Tetovo, less so in the Kumanovo district (marked as Macedonian_and_Albanian in the legend to the map). In the Tetovo district the contact between Albanians and Macedonians is more pronounced since they share a common religion (Islam), whereas in Kumanovo, Macedonians are orthodox, which minimizes inter-ethnic contacts and practically excludes mixed marriages. A similar ethnic distribution can be observed in the earliest reliable source on demographic situation in North Macedonia (Kănčov 1900).

In the western periphery of North Macedonia, the presence of Albanians in the population is very high. Macedonian-Albanian bilingualism is, in principle, characteristic of all of western Macedonia. However, it is religiously unbalanced. As a rule, Albanians in Macedonia – as an ethnic minority – speak Macedonian with the exception of some very old women in remote traditional villages. Macedonians, on the other hand, speak Albanian only if they are Muslim. For the Orthodox Macedonians, Albanian is not a prestigious language. However, in some villages

14 For the Aromanian examples, we apply the "standard" orthography used in the Aromanian media from Macedonia (Ianachieschi-Vlahu 2001: 5–6).

15 Differential object marker for animate objects.

Fig. 3: Ethnic situation in the locations presenting the nothern varieties of North Macedonia in Vidoeski (1999)



of western Macedonia, e.g. in the districts of Struga and Prespa, no ethnolinguistic zoning can be observed, and bilingualism is characteristic of all ethno-confessional groups: not only does the Muslim population speak both languages, but so does the Orthodox (Makarova 2016). In sum, in contexts allowing for bilingualism, Albanians tend to acquire Macedonian more readily than the other way round, i.e. Albanian influence on Macedonian may indeed be expected to occur.[16]

Albanian-Macedonian bilingualism in particular of the Muslim population in this area may also have encouraged the development of the generalization of *mu* in terms of reducing/simplifying the Macedonian pronominal system, which thereby becomes closer to that of Albanian. However, the pattern is not as stable as one could be tempted to predict from the above observations. This instability might be due to the influence of the Macedonian standard. Evidence for this assumption

16 The role of L2 acquisition, in particular by adult speakers, in contact-induced change has long been noted in the literature. Concerning the Balkans, Gołąb (1997: 15) takes this situation as underlying Aromunian-Slavic contact in the 10th c. In a more general perspective, changes observed in spreading languages have been shown to exhibit trends of 'simplification' (i.e. in terms of feature inventory size) as a result of attracting large numbers of L2 speakers (e.g. Nichols 2016; Sonnenhauser 2020 discusses this for Turkish-Macedonian contact during Ottoman times).

is provided by the situation in the South Slavic dialects in Albania and Greece (colored green in Fig. 1): here, Macedonian standard influence is absent, and the generalization of *mu* seems to be systematic and stable.

Concerning pattern 1, the generalization of *mu* as an indirect object marker in the southern dialects, northern varieties of Greek are dominant. Here, it seems that contact did not play a significant role.[17] The pronominal system of northern Greek dialects displays a certain tendency towards simplification processes, for example the spread of ACC into DAT functions with long pronoun forms, but there are no patterns corresponding to those encountered for *mu* in western and southern Macedonian varieties (Bužarovska 2001). This relates to the socio-political situation which discourages speaking Slavic (see above). In addition, Slavic did not have the prestige to attract L2 acquisition by the Greek speaking population anyway.

In sum, while for the generalization of *mu*, contact with Albanian seems indeed a probable factor in certain regions displaying this pattern, it appears less likely for the generalization of *ga*. Even though the relevance of ancient language contact and centuries-old adstratum relations – which is beyond the scope of this paper – should not be excluded, in this case other factors might have played a more prominent role, in particular internal phonological developments, as shown in section 3.2.

3.2 Internal development: loss of phonological distinctions

As suggested by Vidoeski (1965) and Topolinjska (1994: 76), the generalization of *mu* may have resulted from a phonologically driven syncretism in another domain of the pronominal system, namely the generalization of the ACC.3SG.F form *je* for both dative and accusative feminine, on analogy with *te* (ACC.2SG) and *me* (ACC.1SG). This syncretism can be observed in western peripheral Macedonian dialects. An example is given in (10), where the first occurrence of *je* is the 3SG.DAT.F form of the personal pronoun, used in an internal possessive construction.

(10) *majka je je čeka*
mother DAT.3SG.F ACC.3SG.F wait.PRS.3SG
'Her mother is waiting for her' Vidoeski 1965

The reanalysis of masculine singular *mu* and its generalization as the exclusive exponent of dative relations (11) may have emerged as a corollary of this phonolog-

17 According to Bužarovska (2001), Greek is, apparently, responsible for another convergence feature: marking a direct object with the preposition *na*: *ja pitam na mojata zena* 'I am asking my wife'.

ical syncretism. As a consequence, a clearer formal distinction of different types of arguments is retained. The application of *mu* also for the plural in southern dialects can be taken as evidence of its merely syntactic, i.e. argument-indexing function.

(11) *majka mu je čeka*
mother DAT.3SG ACC.3SG.F wait.PRS.3SG
'Her mother is waiting for her' Vidoeski 1965

Concerning the generalization of *ga*, for which contact is less likely than for the generalization of *mu*, phonological and phonetic factors seem to be involved even more. The relevant developments are mainly rooted in the suppletive inflectional paradigm of the third person pronouns in Macedonian. The nominative forms derive from the stem of the Proto-Slavic demonstrative pronouns **tъ* / **to*/**ta*,[18] whereas both clitic/short and stressed/long non-nominative forms derive from the stem of the personal pronoun **jь*/**je*/**ja*.[19] The form of the short accusative feminine pronoun, which is *ja* in the Macedonian standard, varies throughout the territory of the Balkan Slavic dialect continuum. This variability relates to the fact that the reflexes of the Proto-Slavic form **jǫ* were comparatively early substituted for by the genitive form **jejęˀ* > **jęˀ* (OCS ѧ) in some areas (Koneski 1966: 122). The variation in the reflexes of the nasals throughout the South Slavic continuum and in particular in the zone of intersection of Macedonian, Bulgarian, and Serbian borders (see Sobolev 1998: 98–103) added to the overall degree of variation.

Apparently, in the territory of *ga* generalization (yellow in Map 1), Proto-Slavic **ǫ* displays the characteristic West South Slavic reflex *u*. This accounts for the vowel in the form *gu*. The variant *ga*, which also appears in some northern varieties of Macedonian,[20] may either result from the functional expansion, i.e. generalization of the accusative masculine pronoun *ga* (full form *njega*), or from yet another phonetic development of old **jǫ*, namely with the *a*-reflex of **ǫ*, which is typical for western Macedonian dialects. In the latter case, the initial velar [g] most likely developed from initial [*j*], a development known from other languages as well. Cross-linguistically, an initial [*j*] (either etymological or prothetic) may tend to develop into either voiced ([ɣ']) or voiceless ([*x*]) velar fricative (which could develop further in [g]). The former could be seen at certain stages of the pre-standardized

18 The nominative forms derived from the stem of the pronoun **onъ* occur sometimes in northern dialects adjacent to the Torlak zone, but the exact isogloss dividing the *toj*/*taa*/*toa* pattern from the *on* / *ona*/*ono* one is unknown.

19 The full accusative and dative forms descend from that variant of the stem which was used with prepositions, cf. Old Church Slavonic отъ нѣго 'from him', съ нѷмъ 'with him'.

20 Further to the North, in the Torlak area, variants such as *ǵu* and *đu* can be observed (Sobolev 1998: 313–331).

Russian (Kalenčuk & Kasatkina 2013: 52–53). The latter is attested in the development of the initial [*j*] into [*x*] in Spanish (although there are counterexamples as Sp. *ya* < Lat. *jam* 'already'). The question whether the appearance of the initial [*g*] is due to the phonetic development of the unstable initial [*j*] or results from an analogy with the masculine form *go*/*ga*, i.e. from a generalization of the masculine, as suggested by, e.g., Koneski (1966: 122) and Vidoeski (1965: 39), is impossible to answer; both options seem equally likely. Most probably, we are dealing here with a convergence of language internal causes and their blending in the perception of the actual language users, which may have been supported also by language contact.

In addition, the generalization of pronouns may have been reinforced by yet another development in the morphosyntactic system of Macedonian, as shown in section 3.3.

3.3 Areal tendencies: pronouns as object indexes

Given that pronouns prototypically refer to situationally given or contextually introduced nominal referents, the obsolescence of the canonical agreement feature distinctions in gender and, partially, number, is remarkable from a functional point of view. Presumably, this process was stimulated by particular morphosyntactic developments in the Balkan context in the course of which clitic pronouns assumed additional functional weight. Beyond referring to locutors, addressees and third persons, they started to serve as indexes of the core syntactic relations within a verb phrase.

In the standard Macedonian examples in (12a) and (12b), the pronominal clitics *ja* and *go* unambiguously index the direct objects *sestrata* and *bratot*, respectively, while the pronominal clitic *mu* in (13) indexes the indirect object *bratot*.[21]

21 Note that for the indirect object, phrasal dependent marking by *na* is obligatory. This does not contradict the object indexing function of *mu*, since it disambiguates the local/directional adposition *na* marking adverbial constituents, (i), from core arguments of the verb, (ii):

i *Taka pišuva [na biletot]*
like_this write.PRS.3SG on ticket.M.SG.DEF
'This is written on my ticket.' parasolcorpus.org, accessed Nov 13, 2020

ii *Eve što mu*$_i$ *napiša Crvenkovski [na Gruevski]*$_i$
this what DAT.3SG.M write.PST.3SG Crvenkovsi to Gruevski
'This is what Crvenkovski wrote to Gruevski.'
https://daily.mk/what/1798759?about=true, accessed Nov 13, 2020

(12) a. *Brat-ot* *ja*$_i$ *čeka* *sestra-ta*$_i$
brother-DEF.SG.M ACC.3SG.F wait.PRS.3SG sister-DEF.SG.F
'The brother is waiting for the sister'

b. *Brat-ot*$_i$ *go*$_i$ *čeka* *sestra-ta*
brother-DEF.SG.M ACC.3SG.M wait.PRS.3SG sister-DEF.SG.F
'The sister is waiting for the brother'

(13) *Sestra-ta* *mu*$_i$ *kažuva* *na brat-ot*$_i$
sister-DEF.SG.F DAT.3SG.M say.PRS.3SG to brother-DEF.SG.M
'The sister is saying to the brother'

The additional functional weight on these pronominal elements might have led to or accelerated their delexicalization in terms of the bleaching and eventual loss of the referential features gender and number also in their referring function and the generalization of some forms of the paradigm at the expense of others. Processes along these lines have been observed for the development of subject agreement markers from personal pronouns, which, having lost interpretable features such as gender, ended up "with uninterpretable features in a functional position" (van Gelderen 2011: 501).

Cross-linguistically, object indexing has been observed as being sensitive to contact situations. For Romance varieties, Fischer, Navarro & Vega Vilanova (2019) note the loss of feature distinctions on the pronominal index (2019: 60). Since pronominal object indices and short personal pronouns share the same morphological material, the loss of feature distinctions is likely to extend also to the referential uses of pronouns, as evinced in Macedonian.

Both developments are remarkable for Indo-European languages. They prefer gender distinctions in 3rd person pronouns (pointed out in Siewierska 2013),[22] and object indexing on the verb is infrequent in the contemporary varieties (except for the Romance branch, the Balkan areal group and some other individual instances) and volatile in their history, i.e. a pattern that is well-known in the Indo-European varieties but tends to emerge and disappear at a comparatively fast rate (see Dedio & Widmer forthcoming). The fact that both developments are attested in the Balkans suggests a mutual reinforcement within the single varieties and an eventual stabilization in circumstances of language contact.

22 She refers to independent personal pronouns, i.e. long forms. Since the clitic forms under consideration in this paper may also be used independently, drawing on Siewierska's findings seems justified.

3.4 Interaction of causes

The processes sketched above operate in different linguistic domains (phonology, morphosyntax), apply to different patterns, emerge in different sociocultural embeddings, and converge in leading to pronoun generalization. However, interacting with each other, they create an environment that is favorable for the stabilization of a feature that is marginal from an Indo-European perspective.

The first category of causes relates to the agreement features affected by the generalization process and reflects particular morphophonological changes of the pronominal system in different Macedonian varieties. In the case of the generalization of *ga* affecting gender distinctions for accusative relations in the singular, it is likely that we are dealing with the results of two language-internal factors at the same time. The phonetic development of the initial /j/ of Proto-Slavic **jǫ* into /g/ was analogically supported by the presence of the initial /g/ in the masculine form of the pronoun *ga* (reduced from the full form *njega*). Likewise, in the case of *mu*, the internal development of the ACC.F that has led to the syncretism of *je* (ACC.3SG) and *je* (DAT.3SG) might have fostered the generalization of *mu* into a unique exponent of pronominal dative relations and indirect object indexes.

The second category is language-specific. It relates to the structural changes in Macedonian nominal morphology, i.e. the emergence of differential object indexing and the additional function of object indexes assumed by the pronouns. The contact with neighboring Albanian and large-scale bilingualism might have contributed to the spread and stabilization of this development.

The third category represents a cross-linguistically attested process concerning the phonetic development of initial *j* that may have played a role in the formal development of *ga*, supporting in turn its feature-specific development (see above).

The fourth category of causes concerns the role of language contact. In particular feature-specific processes might have been supported by contact influence from the neighboring Balkan languages, with the Albanian patterns (short ACC.F = short ACC.M; short DAT.F = short DAT.M) playing a central role. Table 4 (p. 100) provides an overview of the possible causes and their interaction.

What we see is that the same surface phenomenon may result from different interactions of causes and that language contact displays different influence depending not only on contact strength and direction, but also depending on the specific language internal conditions. In the case of pronoun generalization, this is part of the explanation for why the generalizations have been restricted to either direct (*ga*) or indirect (*mu*) object marking, but did not develop for both, as the Albanian pattern might suggest.

Instead of positing one specific trigger, it is thus more likely if we assume an interplay of factors mutually supporting and reinforcing each other. This has

Table 4: Multiple causes for pronoun generalization

***ga*/*go* /*gu* generalization**	***mu* generalization**
Feature-specific factors	
– Phonetic instability of initial *j* – Formal analogy with masculine *go*/*ga*	– Keeping the distinction of different object arguments despite the collapse of ACC.F and DAT.F (both with surface form *je*)
Language-specific factors	
– Additional functional weight on personal pronouns by being used as argument indexes on the verb leading to delexicalization/bleaching of features and generalization	– Additional functional weight on personal pronouns by being used as argument indexes on the verb leading to delexicalization/bleaching of features and generalization
Cross-linguistic factors	
– Tendency of instable initial *j* to develop into a voiced ([ɣ']) or voiceless ([*x*]) velar fricative	
Contact-specific factors	
– Influence of the Albanian pattern ACC.3SG.M = ACC.3SG.F in the northwest of the territory in question	– Influence of the Albanian pattern DAT.3SG.M = DAT.3SG.F in the northwest and southwest of the territory in question

long been recognized, e.g. by Graur's (1963) 'Principle of Multiple Etymology' and Joseph's (2013) argument for recognizing 'multiplicity of causes' as playing a role in diachronic processes (see in particular Joseph 1983: 179–212 on the causes of the Balkan infinitive loss) instead of positing one single cause, applying Occam's razor to empirical facts.

4 Small-scale patterns in a larger picture

As pointed out above, the generalization of pronouns may have been correlated with another development in the Balkans, namely the emergence and stabilization of differential object indexing, i.e. multiple object representation. Even though the emergence of this strategy is attested in Indo-European, it is characterized by volatility, marginality and cyclicity, following in many cases (outside the Balkans)

a specific trajectory: increasing, reaching a maximum and decreasing without achieving a stable distribution (Dedio & Widmer forthcoming; Haig 2018 speaks of an 'attractor state'). Based on diachronic data from Spanish and Catalan, and data from varieties outside of Spain, Fischer, Navarro & Vega Vilanova (2019: 60) propose five stages of what they call a clitic doubling development cycle, see the illustration in Table 5.

Table 5: Clitic doubling (CD) development cycle

Stage I	→ no CD
Stage II	→ optional CD with full pronouns
Stage III	→ obligatory CD with full pronouns
	→ optional CD with indirect nominal objects [+anim, +def, +spec]
Stage IV	→ obligatory CD with full pronouns
	→ obligatory CD with indirect nominal objects
	→ spread of CD to direct nominal objects [+anim, +def, +spec]
Stage V	→ generalized CD with all objects including inanimate

The most advanced stage (V) in this cycle, attested in Lima and Andean Spanish, is characterized by the loss of the core nominal agreement categories by the pronominal clitic, as in (14), where *lo* does not agree with the object *las plantas* either in gender or in number. Fischer, Navarro & Vega Vilanova (2019: 60) regard this as the "dissolution of the doubling construction" and interpret *lo* as an "agreement marker".

(14) *Eso también* ***lo*** *mata las plantas.*
that too him.M.SG kill the plant.F.PL
'That too kills the plants.'

The data gathered by Dedio & Widmer (forthcoming) and Haig (2018) suggest the cyclic development sketched in Table 5 to be cross-linguistically valid. Following this pattern, some Indo-European languages have developed multiple object representation and kept it in their morphosyntax, in particular in contact situations (Fischer, Navarro & Vega Vilanova 2019: 70), while others have gained and rapidly lost it at some earlier stage. Since object indexing may be gained (e.g. Québécois French) and lost (Lithuanian) quickly, Dedio & Widmer (forthcoming) conclude that it might be pure coincidence whether object indexing is attested in the particular historical variety under investigation or not.

Concerning the Balkan area, the individual processes of emerging object indexing seem to have been synchronized and thus mutually stabilized in their more recent history. The processes of standardization happened to begin at that point of

the history of the core languages of the Balkan area at which multiple object representation – typically a short-lived structure – was on the top of its developmental cycle and has reached a stable distribution. Due to the particular multilingual, small-scale contact situations, the development cycles were synchronized in Albanian, Macedonian and Aromanian before the outset of the various standardization and codification processes. This facilitated the simultaneous existence of this phenomenon in these varieties and their corresponding standards. The stabilization of multirepesentation holds in particular for indirect objects, while for direct objects, it is still subject to semantic and discourse-pragmatic factors. This fits the observation on pronominal feature generalization, which is more stable for dative (*mu*) than for accusative (*ga*) relations.

The observations on the generalization of *mu* and *ga* for indirect and direct object reference and indexing discussed in this paper are valid for Vidoeski's data from the mid-20th century. The present-day situation is different. According to data gathered from our native consultants from the northern part of North Macedonia, the use of *ga* for feminine referents, i.e. generalization pattern 2 for direct objects, is not encountered anymore. Instead, the standard Macedonian form *ja* is widespread, both in referential use and as object index. At the same time, the phenomenon of *mu*-generalization, i.e. generalization concerning indirect objects, still occurs in western varieties, even in the speech of younger informants.

The Macedonian data discussed here amend the development cycle proposed by Fischer, Navarro & Vega Vilanova (2019) by one further stage. This additional stage VI captures the generalization and petrification of one pronominal form and its merger into either a verbal affix or, in case of a certain amount of prosodic freedom, into a specialized agreement particle. In this regard, both northern and southern dialects of North Macedonia (at the moment of their documentation in our data basis, i.e. the Vidoeski corpus) may be at the proposed stage VI[23] of the cycle. At this stage, the pronominal clitic with the indexing functions loses gender and number distinctions and develops into a syntactic agreement marker. The impact of this process is so strong that the generalized form may also be used in purely referential contexts.

Considering the areal and cross-linguistic tendencies, the processes of the pronoun generalization discussed in this paper can be regarded as correlating with an ultimate (but also potentially short-lived) stage of the stabilization of volatile

23 Concerning the northern dialects, this conclusion is only reliable for the time period the texts in Vidoeski's collection were gathered, since we do not yet have modern dialect data from the region. As for the generalization of *mu*, our recent fieldwork data provide regular examples of the feature.

patterns of multiple object representation, triggered by the synchronization of the development cycles in the languages of the area.

5 Conclusion

When a language loses or reduces inflectional case marking, personal pronouns are often the last elements to keep the formal inflectional distinction between object and subject roles (Plungjan 2011: 116). Displaying a considerable reduction of inflectional marking on dependent nominals but not on pronouns, the Balkan Slavic varieties are a typical example for this trend (Escher 2021).

The processes described in this paper suggest that in the Balkan context, the formal distinction of syntactic roles appears to be more stable in the pronominal system than the gender and number distinctions. In the case of the extreme loss of inflectional marking, the marking of the agreement features (gender and number) is lost faster than the formal argument distinction. If we consider the language-internal cause for the generalization of *mu*, we see that what remained after the merger of ACC.F.SG (*je*) and DAT.F.SG (*je*) was the distinction between DAT and ACC, and not that between F and M. In interplay with other factors, this led to the generalization of the masculine form. In the case of the *ga*-generalization in the northern dialects, a phonetic process was the main factor, the impact of which was stronger than the 'need' to keep the canonical agreement features. However, again, the distinction between the core syntactic functions (IO and DO) is kept.

Acknowledgment: The research for this article was supported by the Swiss National Science Foundation (SNSF) grant Nr. 100015_176378/1. This support is gratefully acknowledged.

Abbreviations

Census (2002). "Census of population, households and dwellings in the Republic of Macedonia". Skopje: Republic of Macedonia State Statistica.

Bibliography

Belić, Aleksandar (1905). *Dijalekti istočne i južne Srbije*. Beograd: sine nomine.

Buchholz, Oda von & Wilfried Fiedler (1987). *Albanische Grammatik*. Leipzig: Verlag Enzyklopädie.

Bužarovska, Eleni (2001). "Accusative and dative clitics in southern Macedonian and northern Greek dialects". In: *Balkanistica* 14, 1–17.

Cepenkov, Marko (1892). *Sbornik za narodni umotvorenija, kn. VIII*. Sofia: Nauka i knižnina.

Cheng, Joe, Bhaskar Karambelkar & Yihui Xie (2020). *leaflet. Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.0.2*. URL: https://CRAN.R-project.org/package=leaflet.

Dedio, Stefan & Paul Widmer (forthcoming). "Why dodos matter. The value of unsuccessful variants for the study of language change". In: *Language change in English and beyond. Linguistic theory and historical corpora*. Ed. by Nikolaos Lavidas. Leiden: Brill.

Escher, Anastasia (2021). "Double argument marking in Timok dialect texts (in Balkan Slavic context)". In: *Zeitschrift für Slawistik* 66.1, 61–90. DOI: 10.1515/slaw-2021-0004.

Fischer, Susann, Mario Navarro & Jorge Vega Vilanova (2019). "The clitic doubling parameter". In: *Cycles in language change*. Ed. by Miriam Bouzouita, Anne Breitbarth, Lieven Danckaert & Elisabeth Witzenhausen. Oxford: Oxford University Press, 52–70.

Gołąb, Zbigniew (1997). "The ethnic background and internal linguistic mechanism of the so-called Balkanization of Macedonian". In: *Balkanistica* 10, 13–19.

Graur, Alexandru (1963). *Evoluția limbii române. Privire sintetică*. București: Editura Științifică.

Haig, Geoffrey (2018). "The grammaticalization of object pronouns. Why differential object indexing is an attractor state". In: *Linguistics* 56.4, 781–818.

Ianachieschi-Vlahu, Iancu (2001). *Gramaticā armāneascā*. Crushuva: Drushtvo za kultura i umetnost ART-KUL Krushevo.

Joseph, Brian (1983). *The synchrony and diachrony of the Balkan infinitive. A study in areal, general, and historical linguistics*. Cambridge: Cambridge University Press.

Joseph, Brian (2013). "Multiple sources and multiple causes multiply explored". In: *Studies in Language* 37.3, 675–691.

Kahle, David & Hadley Wickham (2013). "ggmap: Spatial Visualization with ggplot2". In: *The R Journal* 5.1, 144–161. URL: http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

Kalenčuk, Marija L. & Rozalija F. Kasatkina (2013). *Russkaya fonetika v razvitii: fonetičeskie "ottzy" i "deti" nachala XXI veka*. Moskva: Jazyki slavjanskoi kult'ury.

Kănčov, Vasil (1900). *Makedonija. Etnografija i statistika*. Sofia: Balgarskoto knižovno družestvo.

Koneski, Blaže (1966). *Gramatika na makedonskiot literaturen jazik*. Skopje.

Makarova, Anastasia (2016). "Neka zapažanja o etnojezičkoj situaciji u dvojezičnim makedonsko-albanskim selima u regionu Prespa: fenomen Arvati". In: *Studia Slavica Academiae Scientiarum Hungaricae* 61.1, 1–16.

Narumov, Boris P. (2001). "Arumynskij jazyk/dialekt". In: *Jazyki mira. Romanskie jazyki*. Ed. by Irina I. Čelyševa. Moskva: Academia, 636–655.

Newmark, Leonard, Philip Hubbard & Peter R. Prifti (1982). *Standard Albanian. A reference grammar for students*. Stanford, CA: Stanford University Press.

Nichols, Johanna (2016). "Complex edges, transparent frontiers. Grammatical complexity and language spreads". In: *Complexity, Isolation, and Variation*. Ed. by Raffaela Baechler & Guido Seiler. Berlin & Boston: de Gruyter, 117–137.

Plotnikova, Anna A. (2018). "Serbsko-makedonskoe pogranič'e. Osobennosti regiona v kontekste balkanoslavyanskix tradicij (narodnyj kalendar' i sel'skoxozjajstvennaja obrjadnost'". In: *Tradicionnaja kul'tura* XIX.3, 56–67.

Plungjan, Vladimir A. (2011). *Vvedenie v grammatičeskuju semantiku: grammatičeskie značenija i grammatičeskie sistemy jazykov mira*. Moscow: RGGU.

Schenker, Alexander M. (1993). "Proto-Slavonic". In: *The Slavonic languages*. Ed. by Bernard Comrie & Greville G. Corbett. London & New York: Routledge, 60–121.

Seliščev, Afanasii M. (1918). *Očerki po makedonskoi dialektologii*. Kazan': Lito-Tipografija tovariščestva "Umid".

Siewierska, Anna (2013). *Gender distinctions in independent personal pronouns*. Ed. by Matthew Dryer & Martin Haspelmath. URL: https://wals.info/chapter/44.

Sobolev, Andrej N. (1998). "O dialektologičeskom atlase vostočnoj Serbii i zapadnoj Bolgarii". In: *Issledovanija po slavjanskoj dialektologii*. Vol. 5, 106–167.

Sonnenhauser, Barbara (2020). "The virtue of imperfection. Gjorgji Pulevski's Macedonian–Albanian–Turkish dictionary (1875) as a window into historical multilingualism in the Ottoman Balkans". In: *Journal of Historical Sociolinguistics* 6.1, 1–29.

Topolinjska, Zuzanna (1994). *Makedonskite dijalekti vo Egejska Makedonija*. Vol. 1. Skopje: Makedonska akademija na naukite i umetnostite.

van Gelderen, Elly (2011). *The Linguistic Cycle. Language Change and the Language Faculty*. Oxford: Oxford University Press.

Vidoeski, Božidar (1965). "Zamenskite formi vo makedonskite dijalekti". In: *Makedonski jazik* 15, 25–73.

Vidoeski, Božidar (1999). *Dijalektite na makedonskiot jazik*. 2 vols. Skopje: MANU.