

## The Grammaticalization of the Future Tense Auxiliary in the Balkan Languages

Christina Kramer

The formation and varied uses of the future tenses in various Balkan languages raise many theoretical questions pertaining to the diachronic and synchronic development of the verbal categories of future tense and mood; the future tense is also a nice entry point into discussion of Balkan features in general since it is in many respects an understudied point of contrast. Since Sandfeld, the use of the verb *to want* as a future tense auxiliary verb has been included amongst the canonical Balkan features. The varied subsequent developments of the auxiliary *to want* in the Balkan languages and its implications for language typology, language contact, and theories of grammaticalization will be the topic of this paper. Clearly a paper of this length can but lay the groundwork for a longer study. My goal here, therefore, is to present an overview of the problem, some preliminary data, and an outline for a more in-depth study.

In discussions of the future tense in the Balkan languages few studies have provided a comparative overview. Asenova, Belyavski-Frank, Gořab, Ilievski, and most recently Fiedlar are noted exceptions. Most general works on the Balkan Sprachbund cite the shared development of a future formed with the verb *to want* without further elucidation of parallel and diverse developments. When detailed comparisons are made, it is generally from the point of view of morphological development. What is interesting from the typological point of view is that all the Balkan languages developed a future tense with the verb *to want* and a secondary future with the verb *to have* which expresses in different Balkan languages varying degrees of necessity. In addition, none of the languages developed a *go* future such as developed in Romance language and English, though Topolinka (p.c.) suggests that possible examples of a *go* future exist in dialectal Croatian. The subsequent development of these auxiliaries has not been the same in the different Balkan languages. In general it seems that in the southern-most Balkan languages, Albanian, Greek, and Macedonian, the verb *to want* has been fully grammaticalized, while such has not been the case in the other Balkan languages, namely Romanian, Serbian, and Bulgarian, although the process of grammaticalization is evident, particularly in dialects and colloquial speech. Furthermore, Macedonian and the languages in most intense contact with it, Aromanian and Romanian, seem to have undergone parallel developments.

As mentioned above, previous work on the Balkan future has focused on morphological developments and to a much lesser extent on the semantic range of various modal forms. I propose here to analyze these forms on different levels; first, I will employ Traugott's cline of grammaticality, as given in (1):

- (1) Content Item > Grammatical Word > Clitic > Affix

I will also use the paradigm for the grammaticalization of future tenses proposed by Bybee et al., which takes into account the types of modal meaning present in future forms. Bybee et al. (1991: 18–19) propose a path of development of future tenses in which a verb expressing desire can develop future meaning; certain predictions can be made concerning both the types of modal uses of the future and the time at which various uses develop in the

course of grammaticalization.<sup>1</sup> If these theories are correct, then ultimately such a study as proposed here should help determine the relative semantic age of the auxiliary and the degree of grammaticality in each language. In addition such a study may have implications for studies on Sprachbund phenomena, e.g., was the grammaticalization of the future auxiliary a feature borrowed through language contact from one source, or were internal developments in individual languages strengthened due to intense language contact.

A complete study of this phenomenon must determine all the factors outlined below, however since this is a preliminary study, I will comment only on those features which I have been able to analyze from source material. The proposed scale of grammaticality will be based on the following<sup>2</sup>:

- (2) a. Degree of lexical separation of the auxiliary from the lexical verb *to want*
- b. Degree of inflection for person and/or tense in auxiliary and degree of reduction and fusion of the auxiliary verb
- c. Degree of syntactic reinterpretation as clitic rather than verb, as evidenced by presence or absence of main verb preceded by subordinator *të, sa, da, na*, etc.
- d. Syntactic relation to main verb, i.e., can one interpolate other words between the auxiliary and the main verb; e.g., cf. English *I will gladly help*, but \**I will tomorrow help*
- e. Semantic range, i.e., is the future auxiliary used for desire, obligation, probability, prediction, condition, imperatives, iterative/habituals, etc.

Since it appears that the development of the modal verb into a grammaticalized particle spread from south to north I will treat the languages in the following order: Greek, Albanian, Macedonian, Rumanian, Bulgarian, Serbo-Croatian, Romanian. I will discuss Arumanian with Rumanian, even though it is geographically further south.

It has been proposed by Sandfeld, Ilievski, and others that Greek is the source of this Balkan feature. If such were the case we would expect the verb *to want* to be fully grammaticalized in Greek since it is the oldest *want* future. According to the above parameters there are data which give support to this hypothesis. The future particle *tha*, while related to the lexical word *to want*, *thelo* is morphologically and lexically distinct. The auxiliary is no longer transparently derived from the lexical verb and is invariant. The particle is obviously shortened, retaining only the initial phoneme of the verb *thelo* which then fused with the subordinator *na*. The particle *tha* is at the head of the clitic chain and only direct and indirect object clitics can come between the particle and the verb, e.g., *Θα το γράψω* 'I will write it (once)' (Joseph 1987: 173). The fifth criterion is the most difficult to assess, since many published works on different languages do not give a complete, detailed account of varied uses. The particle *tha* cooccurs uninflected with various past and non-past verb tenses and authors of Greek grammars, e.g., Joseph 1987; Householder et al. (1964), note its use in numerous modal contexts. Constructions with *tha* have at least the following meanings:

<sup>1</sup> I will not use here the exact formulation proposed by Bybee et al. The numerical rating they assign to various types of futures may be appropriate at a later stage of this project. They do not account for the development of iterative-habitual meaning and this too will have to be integrated into the current model.

<sup>2</sup> Again, this scale is similar to that proposed by Bybee et al. in that it takes into account formal properties including phonetic shape, syntactic position, separate meanings, but the exact formulation here is my own.

- (3) a. Futurity:  
Θα σου γράφω αύριο.  
'I will write you tomorrow.'
- b. Potential and conditional meaning:  
Θα ειχα χάθει χωρίς εσένα.  
'I would have been lost without you.' [Householder et al. 1964: 105]
- c. Politeness:  
Θα έπρεπε να παλ κανεις να τον φονακσει.  
'Someone should call him.' [Householder et al. 1964: 105]
- d. Probable inference:  
Θα τον έχω γνωρίσει τον κύριο Κόστα.  
'I must have met Mr. Kostas.' [Householder et al. 1964: 105]

Some questions remain to be answered concerning Greek: Is *tha* used for non-modal iterative contexts? What is the difference in degree of modality when *tha* is followed by a perfective or imperfective verb form? Do there exist dialect or colloquial forms in which *tha* can be reduced to an affix, i.e., does the form *th'* exist dialectically.

In Albanian, like Greek, the particle *do* is fully grammaticalized, though a few differences should be mentioned. As in Greek, the particle *do* has become both morphologically and lexically distinct from the verb *to want*, *dua*. It does not inflect for person or tense, but combines with different past and non-past tenses which are subordinated to the particle *të*. It is important to note that dialectally in southern Tosk *do* constructions are formed without *të*, e.g., *do sjell* (Camaj 1984: 153).<sup>3</sup> The only elements which can be interpolated between the auxiliary and the main verb are the direct and indirect object pronouns, which follow *të* and may be fused with it, e.g., *do t'i shkruaj* 'I will write him'. The semantic range covered by constructions formed with the particle *do* includes at least the following:

- (4) a. Future time reference:  
Do të shkoj.  
'I will go.'
- b. Categorical imperatives:  
Ti do të shkosh vetë në Korçë, Demkë!  
'You will go to Korce yourself, Demke!'
- c. Present expectation:  
Ky do të jetë Rrapoja menduan.  
'That will be Rrapoja they thought...'
- d. Future in the past/past probability (with past perfect):  
Ata e kishin parë tek ikte kaluar me tëmën prapa dhe e morën me mend se këtij njeriu diç do t'i kishte ngjarë.  
'They had seen him as he went riding with his mother in the rear and they understood that something must have happened to this man.'

<sup>3</sup> In Geg dialects of Albanian many constructions which in Tosk are formed with *do* are here formed with *kam*. Camaj believes that many constructions formed in Geg with *kam* are giving way to the Tosk standard with *do* (see also Newmark 1982: 86). In all the Balkan languages under consideration here there is a parallel development of a future formed from the verb *to have*. In many of the languages the negative future or emphatic negative future is formed from *to have*. The relationship between Balkan *have* futures and *want* futures will have to be considered carefully in a later study.

- (4) e. Future in the past (with imperfect subjunctive):  
Dulla i tha se atë ditë do të punonin me orar të zgjatur.

'Dulla told them that that day they would work an extended schedule'.

[all examples here are taken from Newmark 1982: 88–89]

Again, numerous questions remain, for example, the degree of acceptability of *do* forms in Geg, and the extent to which *do* can be used in non-modal iterative contexts.

Macedonian shows the highest degree of grammaticalization of *to want* of any of the Balkan languages. In Macedonian the particle *ke* is an invariant particle which forms syntactic constructions with various tense forms. The particle is separated from the lexical word *saka* 'to want'. The particle is closely bound to the verb and only object pronouns can be interpolated between particle and verb. Unlike Albanian, the main verb follows the particle without subordination to the corresponding Macedonian particle *da*, e.g., cf. *do të shkoj* 'I will go' with Macedonian *ke odam*.

In several dialect areas (Vidoeski, ms.) the particle may be further reduced and may occur as the affix *k* before vowel-initial verbs:

- (5) Skopje: E, tamu daleku u tova selo ima edna lamn'a, *k'-idam* da ja tepam.  
Gorno Sonje: Ajde spremaj, *k'-odime*.

The invariant particle *ke* combines with past and non-past tenses and can have at least the following uses:

- (6) a. Future:  
Ke doјdam во седум.  
'I will come at seven.'
- b. Command:  
Ke mi го кажеш крајот на филмот.  
'You will tell me the end of the film.'
- c. Conditionals:  
Da знаев, ke doјdev.  
'If I had known I would have come.'
- d. Iterative/habitual:  
Koga ke се налути не се шегува.  
'Whenever he gets angry he doesn't kid around.'

On the basis of this preliminary analysis, Macedonian, of all the Balkan languages, seems to have the most grammaticalized particle. It has the greatest semantic range and has been the most reduced phonologically and morphologically. Languages in closest contact with Macedonian, in particular Romany and Arumanian, seem to have undergone similar developments; in those languages as well the future auxiliary has become an invariant particle with the broadest range of meanings. While much work remains to be done, the spread of the future auxiliary to iterative/habitual contexts indicates that these forms have the greatest semantic range. Arumanian will be discussed below after Rumanian. Here I will give some very preliminary observations on the Romany spoken in Macedonia.

The available description of Romany as spoken in Macedonia does not provide enough data to answer many of the questions posed here, but from the examples gleaned

from the grammar, the following generalizations can be made. The future is formed from an invariant particle *ka* plus the present tense verb, e.g., *me perava*; *ov perela* 'I fall; he falls'; *me ka perava*; *ov ka perela* 'I will fall; he will fall' (Kepeski and Jusuf 1980: 114). The future-in-the-past is formed with the invariant particle *ka* plus the past tense of the verb followed by the invariant particle *sine*, e.g., *me ka gjiljava sine*; *ov ka gjiljava sine* 'I would have washed; he would have washed'.

Syntactically, it appears that the particle and verb are closely bound and nothing can be placed between them. For example, in the following sentences the Macedonian direct object pronoun *te* 'you' is placed between particle and verb, but in Romany the pronoun *tut* follows the main verb: Macedonian *Kade u da se skrivesh ke te najdam*; Romany *Kote i te garavetu me ka arakhav tut* 'Wherever you hide, I will find you'. The future particle *ka* occurs, in addition to the future and the future-in-the-past shown above, in at least the following types of constructions:

- (7) a. Conditional:  
Te aveva sine manca kji diz, ka nakjhea sine sukar.  
'If you had gone with me to town you would have had a good time.'
- b. Iterative/habitual:  
A ka khuvel o Xhasani, a ka dolen pes ko cingara.  
'As soon as Hasan comes in a fight starts.'

[Kepeski and Jusuf 1980: 116]

In Bulgarian we find a split development in the grammaticalization of *to want*. In non-past contexts in the standard language the verb has developed into a fully grammaticalized particle. It is distinct from the verb *iska* 'to want', as it has been fully deparadigmaticized and occurs at the head of the clitic chain, e.g.:

- |     |                             |   |
|-----|-----------------------------|---|
| (8) | Future:                     | Future perfect:                               |
| a.  | Ще каза, ще кажеш           | c. Ще сум/бъда носил, ще си бъде носил.       |
|     | 'I will say, you will say.' | 'I will have brought, you will have brought.' |
| b.  | Ще го видиш ли?             |   |
|     | 'Do you see him?'           |   |

In past tense forms, however, *ще* remains a verbal auxiliary. In constructions with the so-called future-in-the-past the auxiliary is conjugated, it is not syntactically bound to the main verb, and the main verb is subordinated to the particle *да*, e.g.:

- (9) a. Щях да пиша; щеше да пишеш.  
'I would have written, you would have written.'
- b. Щях да съм/да бъда ходил; щеше да си/да бъдеш ходил.  
'I would have been gone, you would have been gone.'
- c. Детето щеше още вчера да си е написало домашните упражнения.  
'The children would have already completed their homework yesterday.'

There is a tendency towards grammaticalization of the verbal auxiliary in at least some of the past tense forms, note for example the variant forms of the past reported: *щях съм да съм донесъл/ще съм донесъл*.

Constructions with *ue* cover a wide range of meanings, including future, conditional, attenuated commands, presumption, and iterative/habituals. Thus, at the semantic level the auxiliary is widely developed, but syntactically it is conservative. There is clearly a tendency in the colloquial language towards the development of an invariant particle. Bulgarian, seen from the context of Balkan Slavic, is clearly transitional between Serbo-Croatian and Macedonian in the development of an invariant particle.

If we look at the fate of *to want* in Serbo-Croatian, it is clear that the verb *to want* has undergone some of the same processes as in Greek and Albanian, but that in the verbal system *to want* plays a much narrower role. Belyavski-Frank (1984) has pointed out that *to be* is here the primary modal auxiliary and is used, for example, for hypothetical conditionals and past iterative/habituals. In literary Serbo-Croatian and most of the dialects, *to want* as a modal is limited to its enclitic forms used for the future. According to the above criteria, the future formant has undergone less grammaticalization compared with Greek, Albanian, and other Balkan Slavic: Macedonian and Bulgarian. First, in its full form the auxiliary is homonymous with the verb *to want*. Syntactically, various other elements can come between the auxiliary and the main verb. In fact, there are different syntactic rules here: the auxiliary must stand in second position, but the main verb is not bound syntactically to the auxiliary, e.g.,

- (10) Dosta će stara biti ova poslovice.  
'This proverb will be old enough.'

Belyavski-Frank cites the following from various Serbo-Croatian dialects, where constructions with *want* have a wider possible range of meanings:

- (11) a. An event which nearly happened:  
Htedoh juće da poginem, pukla mi je guma.  
'I almost got killed yesterday, I had a blowout.'  
b. An event which is about to happen; an inevitable, involuntary action:  
Kad je već pod starost htio da umre...  
'When he was already very old and on the point of death...'

In most of the southern dialects, according to Belyavski-Frank, constructions with *want* still have only limited semantic functions and are in competition with constructions with *bi*. In constructions with the imperfect of *to want* followed by the infinitive several meanings are possible. In the Southwestern dialects of Montenegro, *htěti* is most highly developed as a modal auxiliary and can be used in past conditions, iterative/habitual, and attenuated expressions. It is clear, then, that the verb *to want* according to our above parameters has undergone some processes towards becoming grammaticalized, but much less so than in the southern Balkan languages. The picture that Belyavski-Frank draws, however, is one of wide dialectal variation, which will have to be taken into account in a broader study of the grammaticalization of *htěti*.

Romanian has four ways of forming the future: at least two are formed from *voi* 'to want', one from *avea* 'to have' (which will not be treated here), and one which may have developed from either *to want* or *to have*.<sup>4</sup> The three futures formed from *voi* are as follows:

- (12) a. *voi* *vom*  
*vei* *veți*  
*va* *vor*  
b. *o + să*  
c. colloquial and dialectal forms which show loss of initial *vin* type (a) and the alternation of *a/o* in the third-person singular

Types (a) and (b) combine with an invariant short infinitive, e.g., *Eu voi veni* 'I will come', *Tu vei veni* 'You will come', etc. The verb is clearly related to the lexical verb *to want*, even though it is morphologically distinct, cf. *voilvreau*. There is potential ambiguity in the third-person plural, where there is homonymy in the two forms, i.e., *vor* 'they will'/'they want'. However, the two forms behave differently syntactically; cf. *vor veni* 'they will come' vs. *vor să vine* 'they want to come'.

Unlike Greek, Albanian, Macedonian, and Bulgarian, Romanian pronominal clitics precede the auxiliary, e.g., *O vom vedea* 'We will see her'. A limited set of adverbs, however, can be inserted between auxiliary and verb, e.g., *Voi mai veni* 'I shall read again', *Vei și citi* 'You will also read'.

These two future forms have a more restricted range of meanings, but are used at least in contexts with future and conditional meaning. Preliminary research indicates that such constructions are not used for future in the past, past conditionals, present supposition, or iterative/habituals. Thus, according to our cline of grammaticality, these forms are less grammaticalized. The auxiliary is more verbal, more complex morphologically, less bound syntactically to the verb, and has more restricted range of meanings.

Unlike the futures formed with the conjugated auxiliaries, which are considered to be somewhat bookish and are chiefly written forms (Cazacu et al 1982: 348), the invariant form of the auxiliary *o* plus *să* is widespread in the vernacular. This auxiliary looks much more like the canonical Balkan future form. Syntactically it behaves like *do* in Albanian: it is an invariant auxiliary; it is distinct from the lexical verb *to want*; it is followed by the subjunctive; and object clitics either fuse to the subordinating particle, if vowel-initial, or come between subjunctive particle and the main verb. There is a limited set of adverbs, however, which can be placed between *o să* and the main verb, including *mai* 'again', *și* 'also', *cam* 'almost', *prea* 'too' and *tot* 'still'; e.g., *O să mai vin* 'I will come again', *O să le-o mai și spui* 'You will also tell it to them'.

The future auxiliary in Arumanian must also be considered. Again, there are at least two separate developments. First, the fixed form *vrea* occurs in constructions with the conditional, the imperfect subjunctive, and the present subjunctive. It is invariant, though transparently related to the verb *to want*. It has a wide range of semantic functions, including the past conditional and expressions of the type 'X nearly happened,' or 'to be on the point of an involuntary action' (Belyavski-Frank 1984). A newer construction is formed with the particle *va*, as described by Goțab in his work on Krușevo. This particle is invariant and less transparently related to the verb *to want*. Like Albanian *do të* and Rumanian *o să*, it forms analogous syntactic constructions of modal auxiliary followed by the subordinating conjunction *-s-* plus main verb, e.g., *Va-s-videam*, *va-s-videai*. The main verb can only be separated from the auxiliary by the object clitics. The subordinator *-s-* is often not used, however, if the object clitics are present, e.g., *va-l-lu-akumpr* 'I will buy it for him'. The range of meanings with this construction is broad, including at least future, future-in-the-past, conditional, and iterative/habituals.

<sup>4</sup> See Halvorsen 1973 for a description of the differences in usage of the separate futures.

Let us now place Greek, Albanian, Macedonian, Bulgarian, Serbo-Croatian and Romanian along a line showing a continuum from minimum to maximum presence of a given feature. Arumanian and Romany are not included, since the least data has been collected on them. We see a pattern again develop of the least grammaticalized future form in the north to the most grammaticalized in the south.<sup>5</sup>

- (13) a. Reduction of lexical word to invariant particle:  
SC > Rom > Bg > Alb/Gr/Mac
- b. Presence of subordinator *da/tē/sā/na* between particle and verb:  
SC > Rom > Bg (+past) > Alb > B (-past)/Mac<sup>6</sup>
- c. Degree of bound syntactic position with main verb, i.e., whether words can be interpolated between particle and verb:  
SC > Rom > Bg (+past) > Bg (past)/Alb/Gr/Mac
- d. Scope of semantic range, i.e., future, conditional, command, iterative/habituals, etc.  
SC > Rom > Alb/Bg/Gr > Mac

This model correlates in many respects with the time line of development given in Asenova (1989) and the model of grammaticalization proposed at the beginning of this paper. Asenova graphed the historic development from lexical word to particle. She gives evidence that the verb *to want* developed into an invariant particle in Greek and Bulgarian before it did in Romanian. Her timeline shows a later development for Albanian. Future work will have to account for the split development in Bulgarian, the multiple developments in Romanian, the high degree of grammaticalization in the Albanian form if it was indeed a later development, and many other similar questions.

In addition, a more complete study on the grammaticalization of the future forms from *to want* in the Balkan languages will contribute to discussions concerning the nature of clitics, particles and words. The work by Zwicky (1984) will be particularly relevant. Too often a comparison of Balkan features does not take into account different roles a form plays within a separate linguistic system. For example, Macedonian *ke* is a particle, but it is not clear what the difference is in the Bulgarian forms *šte* and *šteše da*. Furthermore, this study will add to studies on the interrelationship between future forms and modality. There is much work to be done not only in the diachronic development and synchronic description of these future constructions in the standard languages of the Balkans, but also in the mapping of these features in the dialects as well.

<sup>5</sup> The following abbreviations have been adopted:

Alb	Albanian	Gr	Greek	Rom	Romanian
Bg	Bulgarian	Mac	Macedonian	SC	Serbo-Croatian

The slash / is used to designate parallel development.

<sup>6</sup> Greek could be considered here with Bulgarian and Macedonian, since synchronically the subordinator *na* has fused with *thelo* and is not longer felt to be a separate morpheme in these constructions.

## References

- Asenova, Petia. (1989) *Balkansko ezikoznanie: osnovi problemi na balkanskija ezikov sâjuz*. Sofia: Nauka i izkustvo.
- Belyavski-Frank, Masha. (1984) "On the status of three modal auxiliaries in Balkan Slavic and Romance". *Papers for the V. Congress of Southeast European Studies*. Columbus: Slavica.
- Bybee, Joan, William Pagliuca, and Revere D. Perkins. (1991) "Back to the future". Elizabeth Traugott and Bernd Heine, eds. *Approaches to grammaticalization*. Amsterdam: John Benjamins, v. 2.
- Camaj, Martin. (1984) *Albanian grammar*. Wiesbaden: Otto Harrassowitz.
- Cazacu, Boris et al. (1982) *A course in contemporary Romanian*. Bucharest: Editura Didactică și pedagogică, 3rd ed.
- Fiedler, Wilfried. (1989) "Zur Arealtypologie der Futurbildung in den Balkansprachen". *Linguistische Studien* 192. Berlin: Akademie der Wissenschaften der DDR.
- Goľab, Zbigniew. (1984) *The Arumanian dialect of Kruševu in SR Macedonia*. Skopje: Macedonian Academy of Sciences and Arts.
- Gramatika na sâvremennija Bâlgarski knižoven ezik, II: Morfologija*. Sofia: Balgarski Akademija na naukite, 1983.
- Gramatika limbii Romine*. Bucharest: Academia Republicii populare Romine, 1963.
- Heine, Bernd, Ulrike Claudi, and Friederike Hunnemeyer. (1991) *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.
- Householder, Fred W., Kostas Kazazis, and Andreas Koutsoudas. (1964) *Reference grammar of literary Dhimotiki*. Bloomington: Indiana University Press.
- Ilievski, Petar. (1988) *Balkanološki lingvistički studii*. Skopje: Institut za makedonski jazik.
- Joseph, Brian D and Irene Philippaki-Warbuton. (1987) *Modern Greek*. London: Croom Helm.
- Kepeski, Krume and Šaip Jusuf. (1980) *Romani gramatika*. Skopje: Naša kniga.
- Newmark, Leonard, Philip Hubbard, Peter Prifti. (1982) *Standard Albanian: A reference grammar for students*. Stanford: Stanford University Press.
- Sandfeld, Kr. (1930) *Linguistique Balkanique*. Paris: Librairie C. Klincksieck.
- Stevanović, M. (1971) *Gramatika srpskohrvatskog jezika*. Cetinje: Obod.
- Traugott, Elizabeth Closs and Bernd Heine. (1991) *Approaches to grammaticalization*. Amsterdam: John Benjamins, v. 2.
- Vidoeski, Božo. Unpublished manuscript of dialect materials.
- Zwicky, Arnold M. (1985) "Clitics and particles". *Language* 61: 2.

Department of Slavic Languages  
University of Toronto  
21 Sussex Avenue  
Toronto, Ontario M5S 1A1 Canada  
kramer@epas.utoronto.ca