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. Aserova, Belyavski-Frank, Golab, Hlevski, 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 Topolinska (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 Romany, 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. 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 grammaticalization 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:

(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 τε, sa, do, 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/habituative, 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, Romany, Bulgarian, Serbo-Croatian, Romanian. I will discuss Arumanian with Rumanian, even though it is geographically further south.

It has been proposed by Sandfield, 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 thea 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 to γιορτάσω '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/coocurs 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:

(3) a. Futurity:
    Θα σου γράφω αύριο.
    'I will write you tomorrow.'

b. Potential and conditional meaning:
    Θα έπρεπε να πας να τον βοηθήσει.
    'I would have been lost without you.' [Householder et al. 1964: 105]
c. Politeness:
    Θα έπρεπε να πάγαγε κάνεις να τον βοηθήσεις.
    'Someone should call him.' [Householder et al. 1964: 105]
d. Probable inference:
    Θα τον έφαγες τον καρπό του Κοστα.
    '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 τα 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 τε. It is important to note that dialectally in southern Tosk do constructions are formed without τε, e.g., do sjet (Carnaj 1984: 155). The only elements which can be interpolated between the auxiliary and the main verb are the direct and indirect object pronouns, which follow τε and may be fused with it, e.g., do ti shqyr '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 dukshet vetë më Korçë, Demkë!
    'You will go to Korce yourself, Demke!'

c. Present expectation:
    Ki do të jetë Rrapoja menduan.
    'That will be Rrapoja they thought...'

d. Future in the past (probability with past perfect):
    Ata e kishin parë tek ikte kaltar me témnin prapa dhe e morën me mend skëtit njërit diç do të 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.'

In Gag dialects of Albanian many constructions which in Tosk are formed with do are here formed with kum. Carnaj believes that many constructions formed in Gag with kum 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 tsa atse ditte do te vunon mba orar te zsigat.
    ‘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 te
siga: ‘I will go’ with Macedonian ke odam.

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

(5) Skopje: E, tamu dalke u tova selo ima edda larnu, k-’idam da ja tepeam.
    Gorno Sorje: 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 dodam vo sedam.
    ‘I will come at seven.’

b. Command:
    Ke mi go kajsho krajot na filmot.
    ‘You will tell me the end of the film.’

c. Conditionals:
    Da zogov, ke dozdr.
    ‘If I had known I would have come.’

d. Iterative/habitual:
    Kogu ke se naklu ne se shiguvka.
    ‘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 Aromanian, 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. Aromanian 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 Jusu 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 gjithva sine; ov ka gjithvaja 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 tu
follows the main verb: Macedonian Kade u da te eipeuke ke me nakhem; Romany: Kote i te
garvuce te ke arakhv tu ‘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 ake sine manca kji diz, ka nakjesa sine sukar.
    ‘If you had gone with me to town you would have had a good time.’

b. Iterative/habitual:
    A ka sahu o Xhasani, a ka dole pes ko cingara.
    ‘As soon as Hasan comes in a fight starts.’
    [Kepeski and Jusu 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
paradigmmatized and occurs at the head of the clitic chain, e.g.,

(8) Future:
    a. Sce kaz, sce kajesh.
       ‘I will say, you will say.’
    b. Sce ssumblja nosilja, sce sibilja nosilja.
       ‘I will have brought, you will have brought.’

In past tense forms, however, the verb 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 da, e.g.,

(9) a. Zhe da pisca; zhece da pisipat.
    ‘I would have written, you would have written.’

b. Zhe da ssumblja bila holja; zhece da sili da bila holja.
    ‘I would have been gone, you would have been gone.’

c. Deteta zhece vjera da ci ese naipalo domashniye upravleniya.
    ‘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: zhe
sik da sili donesla/ dzie sili donesla.
Constructions with *voi* cover a wide range of meanings, including future, conditional, attenuated commands, presumption, and iterative/habitudinal. 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 *be* is here the primary modal auxiliary and is used, for example, for hypothetical conditionals and past iterative/habitudinal. In literary Serbo-Croatian and most of the dialects, *to want* as a modal is limited to its elicitic 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 *de stara hiti* ova poslovica.
    This proverb will be old enough.

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

11. a. An event which nearly happened:
    Hrodhju oju o Zagrebu, pokla mi je guma.
    "I almost got killed yesterday, I had a blowout."

b. An event which is about to happen; an inevitable, involuntary action:
    Kadi je vec pod starost hito da mame...
    "When he was already very old and on the point of death..."

In most of the southern dialects, according to Belyavski-Frank, constructions with *voi* 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, *hiti* is most highly developed as a modal auxiliary and can be used in past conditions, iterative/habitudinal, 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 *hiti*.

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

<table>
<thead>
<tr>
<th>(12)</th>
<th>a. <em>voi</em></th>
<th><em>vor</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>b. <em>o</em></td>
<td><em>să</em></td>
<td></td>
</tr>
<tr>
<td>c. colloquial and dialectal forms which show loss of initial <em>vor</em> type (a) and the alternation of <em>o</em> in the third-person singular</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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. *voi/vereau*. 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* "they will come" vs. *vor* să *vin* "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*. *Vezi și citește* *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 suposition, or iterative/habitudinal. Thus, according to the case 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 al et al. 1987: 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 verbs, including mai *again*, și *also*, cum *almost*, prea *too* and tor *still*; e.g., *O să mai văd* *I will come again*. *O să le-o mai și găsea* *You will also tell it to them*.

The future auxiliary in Romanian must also be considered. Again, there are at least two separate developments. First, the fixed form *verea* 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 "that nearly happened." Or to be on the point of an involuntary action" (Belyavsky-Frank 1984). A newer construction is formed with the particle *va* as described by Golab in his work on Krusevo. This particle is invariant and less transparently related to the verb to *want*. Like Albanian *do* *te* and Romanian *o* *să*, it forms analogous syntactic constructions of modal auxiliary followed by the subordinating conjunction *-s* plus main verb, e.g., *Va-ti vinde* *We 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.
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. Armenian 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.  

(13) a. Reduction of lexical word to invariant particle:  
SC > Rom > Bg > Alb/Gr/Mac  
b. Presence of subordinator dâste să na between particle and verb:  
SC > Rom > Bg (+past) > Alb > B (>past)/Mac  
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/habitual, etc.:  
SC > Rom > Alb/Gr/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 Rumanian, 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 ho is a particle, but it is not clear what the difference is in the Bulgarian forms ho and stve 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.

---

References


Videvski, Božo. Unpublished manuscript of dialect materials.


---

The following abbreviations have been adopted:

<table>
<thead>
<tr>
<th>Alb</th>
<th>Albanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bg</td>
<td>Bulgarian</td>
</tr>
<tr>
<td>Cr</td>
<td>Greek</td>
</tr>
<tr>
<td>Mac</td>
<td>Macedonian</td>
</tr>
<tr>
<td>SC</td>
<td>Serbo-Croatian</td>
</tr>
<tr>
<td>Rom</td>
<td>Romanian</td>
</tr>
</tbody>
</table>

The slash is used to designate parallel development.

Greek could be considered here with Bulgarian and Macedonian, since synchronically the subordinator na has fused with theo and is not longer felt to be a separate morpheme in these constructions.