

# Modality in Turkic between arbitrariness, grammaticality and lexicality

Julian Rentzsch\*

## Proem

The following discussion on certain aspects of the expression of modality is illustrated with data from the Turkic languages, but it should basically apply to many non-Turkic languages as well.

Modality is primarily a semantic term. It is used as an umbrella term to cover a number of subclasses such as alethic, epistemic, dynamic and deontic modality. It is a matter of dispute in the literature how many and which classes of modality are linguistically relevant and how far such categories as mood, interrogation, negation and evidentiality should be subsumed under the category of modality. Work that has been influential in the linguistic discourse on modality and the formation of terminology includes Lyons (1977), Palmer (1986 [2001]), Bybee & Perkins & Pagliuca (1994) etc. (see Nuyts (2006) and de Haan (2006) for surveys). Classificational and terminological issues are not the central object of this paper and will only be touched upon in passing; for my opinion on certain topics, see Rentzsch (2010). In this article, the terms epistemic modality and event modality (the latter taken from Palmer) will be used to distinguish two of the main semantic classes.

As a semantic category, modality can be studied from the perspectives of philosophy and formal logic as well, but the present paper will be confined to linguistic issues. Pre-supposing an understanding of linguistic semantics as the study of form-meaning correspondences, i.e. the study of linguistic items that entail both a formal and a semantic aspect (an understanding that is frequently ignored in the literature), this paper will deal with linguistic material with modal meaning, i.e. expressions of modality that consist of lexical or grammatical items or a combination of both.

The involvement of linguistic material in the expression of modality, i.e. the mere fact that the category has a formal side, has further implications. Lexical items are integrated into clauses and sentences, i.e. into structures with specific syntactic rules. Grammatical markers crave a unit on which to operate, like a noun, a noun phrase, a verb stem, a predicate etc., and combined with this unit they also become part of clauses and sentences. As a result, all linguistic expressions of modality are either subject to the syntactic rules of the respective language or they inaugurate new ones, in any case they are part of syntactic layers and scope hierarchies. Hence, modality is not only a semantic class, but also a functional one.

\* University of Szeged.

Modal classes can be divided roughly by functional criteria according to their relation to the aspect category (i.e. items like Turkish *-(V)yor*, Kazakh *-A žatır* or Uzbek *-yapti*), resulting in three possible classes, viz modality<sub>1</sub> (MOD-1, items potentially in the scope of aspect), modality<sub>2</sub> (MOD-2, items on the same layer as aspect, i.e. interchangeable with and mutually exclusive to aspect) and modality<sub>3</sub> (MOD-3, items that potentially have aspect in their scope). This functional subdivision can easily be tested, and it is not confined to the Turkic languages, but valid for all languages with a salient aspect category. It is not entirely arbitrary (although it could be supplemented by more minute subdivisions) as the three functional classes acquired through the relation to the aspect category display characteristic semantic traits; that is the functional classes are paralleled by semantic distinctions. Modality<sub>1</sub> typically denotes event modality, modality<sub>3</sub> signifies epistemic modality, and modality<sub>2</sub> encompasses a heterogeneous set of categories, including the moods, the future/prospective and the (underived) conditional. The semantic classes are “transparent” to a certain degree in that items of lower ranging functional classes can be interpreted, given an appropriate context, in terms of a higher ranging semantic class (e.g. modality<sub>1</sub> in terms of epistemic modality, etc.). These basically semantic interpretations can develop into actual semantic meanings as a result of grammaticalization processes.

### Arbitrary and conventionalized expressions of modality

Totally arbitrary expressions of modality, i.e. completely spontaneous combinations of lexical and grammatical items entirely void of any preexisting model, are in fact rarely encountered. Most modal expressions are to a certain degree conventionalized, i.e. occur frequently enough that they can be assumed to constitute part of a common heritage of a group of speakers. However, it is immediately obvious that conventionalization can be graded according to a combination of several parameters, like dissemination throughout speaker groups and registers as well as frequency. Establishing the degree of conventionalization is not straightforward, as even highly grammaticalized items can occur very infrequently (e.g. dying forms). The issue becomes more obvious as soon as a change can be observed in form or meaning. Significant formal change, such as phonetic reduction, alteration of a government pattern etc., contradicts arbitrariness (which is the free combination of linguistic material according to the rules of a given language). “Significant” means that the phenomenon may not be confined to just one singular utterance or the speech habit of one individual speaker (while “mistakes” shared by a group of speakers may be significant). Likewise, as soon as the meaning of a given complex entity differs from the combined meaning of its components or even is no longer transparent from a synchronic point of view, we may infer that the expression has undergone a process of conventionalization. On the other hand, there are numerous examples of highly conventionalized expressions that both are semantically transparent and display no trace of formal change. Assessing their degree of conventionalization in a non-intuitive, objectively verifiable way is extremely difficult and calls for the consideration of additional criteria besides frequency and dissemination, e.g. the criterion of paradigmaticity, i.e. the question of whether or not a given unit is part of a set of items that enter more or less systematic oppositions. The problem can be illustrated by the following three examples from



Turkish *that* feature modal constructions with different meanings, none of which displays a change either in form or meaning:

- (1) *Hepimiz bu kelimelerin hikmetini düşünme gereği duymuş olacağız ki kimse-den çıt çıkmadı.*  
‘Probably all of us felt the need to think about the deeper meaning of these words, for nobody made a sound.’  
(Şafak 2009: 300)
- (2) *Şimdi bu adamlar, ülkücü takımından olabilir, milliyetçilik filan. Gençliklerinde bazı olaylara karışmış olmaları da muhtemel.*  
‘Now, these men may be part of the idealists, nationalism and so on. It is also probable that they were involved in some events when they were young.’  
(Ümit 2006: 317)
- (3) *Raporumu hazırlamam için de olay yerini görmem gerek.*  
‘And in order to write my report I have to see the site of the event.’  
(Ümit 2008: 66)

Among the examples cited, the complex unit [A] *-mA gereği duy-* ‘to feel the need to X’ in ex. (1) is clearly situated most closely to the arbitrariness end of the scale. All of its constituents and their combined meaning are totally transparent. The noun *gerek* ‘necessity’ could be replaced by any noun that can meaningfully be combined with a verbal noun and the verb *duy-* ‘to feel’ (such as *ihtiyaç* ‘need’ and *aşk* ‘love’) while the verb *duy-* could be replaced by some other verbs (such as *çık-* ‘to emerge’). On the other hand, the construction is not entirely arbitrary. A Google query (January 10, 2011) for “gereği duydum” rendered more than 200,000 results for this particular combination alone (i.e. with the first person singular of *-DI*), which is a strikingly high number compared to only 8,000 results for “gerek duydum”, which intuitively is a no less meaningful combination. Moreover, *-mA gereği duy-* is fossilized to a certain extent in that *gereği duy-* virtually always combines with a verbal noun (the one in *-mA*). In other words, the segment *X-mA* can hardly be replaced by a plain noun. If a feeling of necessity is to be expressed in respect of a plain noun, the construction is <N-DAT gerek duy->. That is, the noun is mentioned in the dative case, and *gerek* does not take the possessive marker. (This option is even applicable to verbal nouns, i.e. *düşünmeye gerek duy-*. This variant, though, is much less frequent than *-mA gereği duy-* and is actually a candidate for an arbitrary expression.)

The epistemic construction [B] <X-ASP olması muhtemel> featured in ex. (2) is considerably more frequent than the aforementioned expression and not uncommon in both spoken and written registers. However, many Turkish speakers use alternative constructions with a similar meaning much more frequently, such as those involving the adverb *muhtemelen* (*muhtemelen karışmıştır*), <X-ASP olacak> (*karışmış olacak*) or <X-ASP olmalı> (*karışmış olmalı*). (Needless to say, the four options mentioned are not entirely identical in meaning, but similar enough to be interchangeable in many contexts.)

Finally, [C] <-mA-POSS+gerek> as exemplified in (3) is one of the most common expressions of necessity in Turkish (together with <-mA-POSS+lazım> and the variant with the auxiliary verb *gerek-*, <-mA-POSS+gerek->). It is frequent in all registers of spoken and written Turkish and forms part of the standard repertoire of all speakers.

We can be sure that [C] is **situated** quite high on the conventionalization scale by virtue of its mere frequency and wide dissemination among virtually all speaker groups. On the other hand, the rate of occurrence should not be overestimated as a criterion for establishing the degree of conventionalization. Low frequency does not necessarily correspond to a low degree of conventionalization. There are other factors that contribute to the frequency of a linguistic item besides the relative distance covered on a path (or “cline”) that leads away from arbitrariness. [B] encodes a notion completely different from [C], making these constructions difficult to compare in terms of frequency. Both extralinguistic and text-specific factors may contribute to the ratio of usage simply by the fact that one item is less often needed than the other. Construction [A] craves a conscious subject that is capable of “feeling” the necessity while the semantically related construction [C] does not require a conscious subject. This renders the latter eligible for more situations than the former, thus influencing its rate of occurrence. The frequency of a given item may also be affected by the existence of alternative expressions. Several choices in the same situation naturally result in a decrease in incidence for all of them.

Another criterion that distinguishes [B] and [C] on the one hand from [A] on the other is that the former two belong to sets of systematically contrasting items that encode various shades of epistemic modality ([B]) and event modality ([C]), respectively, although these two sets do not represent strictly closed classes. In contrast, [A], while being an expression of necessity, has no immediate parallel construction for possibility or intention and hence lacks the paradigmaticity of [B] and [C]. Paradigmaticity, on the other hand, is one of several criteria of grammaticalization. This implies that [B] and [C] are (in some respects) “more grammatical” than [A].

This brings us to another problem, which has to do with the nature of conventionalization. Clearly, conventionalization is not a unilinear process that leads from an arbitrary starting point to one conventionalized endpoint. In the literature on language evolution, two main tracks of conventionalization are distinguished: grammaticalization and lexicalization, the former resulting in grammatical items, i.e. morphemes or morphosyntactic constructions, the latter producing lexemes and idioms. The process of idiom formation is frequently also labeled idiomaticization, but the development is essentially the same for lexemes and idioms and will be treated together in the present framework.

The relationship between grammaticalization and lexicalization is a complicated issue, as the conception of lexicon and grammar as two separate boxes, one of which contains grammatical items, the other lexical ones, is highly problematic (cf. Himmelmann 2004: 21). Grammar and lexicon are interdependent and in permanent interaction. Virtually every linguistic utterance combines grammatical and lexical items. Lexemes are integrated into constructions, which are subject to the rules of grammar, and grammatical items combine with lexemes. In a diachronic perspective, lexemes can be integrated into morphosyntactic constructions that become conventionalized and become part of the grammatical system of a language. Thus, while it makes sense to distinguish lexicon and grammar on a certain level, they are not strictly discrete categories.

Before attempting to reconcile grammaticalization and lexicalization, we will consider some grammatical and lexical expressions of modality in order to convey a rough idea about possible stages of conventionalization.



## Grammatical expressions of modality

Grammatical expressions are common in all three functional classes of modality, but the most strongly grammaticalized items tend to fall under MOD-2. The aspect/prospective/mood layer seems to be the most integrated layer in the domain of inflectional verb morphology, probably because this “slot” must be filled in every finite predication that has a verbal core (in the Turkic languages). MOD-1 and MOD-3, on the other hand, are always optional and hence less frequently marked, thus slowing down the phonetic attrition, generally speaking. Despite this tendency, there are both younger, non-reduced items in the domain of MOD-2 and quite strongly reduced ones in the domains of MOD-1 and MOD-3.

Examples of synchronically unanalyzable, monomorphemic and hence maximally grammaticalized MOD-2 items are the basic voluntative and imperative items, the conditional in *-sA(r)* and prospective items like *-tAçI*, *-GAy*, *-(y)AsI* and *-(y)AjAK*. A younger item that is clearly grammaticalized, but much less advanced than these on the grammaticalization cline, is the Turkish discontinuous construction *keşke* plus the conditional, with an optional past marker *idi* to add an irrealis component (4). This is an emotive (mood) operator that semantically resembles – depending on the personal marking – voluntative and optative items, with the difference that the renewed construction expresses the desire of the conscious subject more emphatically (“markedly”) than the older, established morphemes. This construction also demonstrates how lexical material, in this case the lexeme *kāški* borrowed from Persian (which in itself has been lexicalized from two components), can be integrated into morphosyntactic constructions that are being grammaticalized.

- (4) *Keşke seni daha önce tanışaydım.*  
 ‘Would that I had met you earlier!’  
 (Şafak 2009: 368)

A commonly observed phenomenon is that of grammatical markers gradually losing their semantic emphasis as they proceed along the grammaticalization cline and being replaced by renewed, semantically more narrow items. Initially, the renewed item covers only some specialized functions within a given semantic domain, leaving the rest of the domain to the older item. As the renewed item broadens its scope within that semantic field, it gradually pushes the old item out of its former domain, causing the latter to fade out or to maintain only some marginal functions.

An example of this process is found in the domain of possibility. The normal marker of (event modal) possibility in Turkish is *-(y)Abil-*, a morpheme resulting from the fusion of the converb *-(y)A* with the auxiliary *bil-* ‘to know’. This marker is capable of covering more or less the whole field of possibility in Turkish. There is, however, a younger item *-mAyI bil-*, which only covers a specific subdomain of possibility, namely internal possibility (cf. van der Auwera & Plungian 1998). As the established unit *-(y)Abil-* is still able to cover this subdomain, there is an opposition [ $\pm$ internal] in Turkish, with *-mAyI bil-* as the marked member and *-(y)Abil-* as the unmarked one (privative opposition). A closer look at the Turkic language history reveals that the opposition [ $\pm$ internal], as well as the process of renewing its marked member, has been present throughout almost the entire documented history of this language family.

The earliest attested Turkic mainstream possibility marker consists of a converb either in  $-(y)V$  or in  $-GAlI$  (which will be labeled CV summarily here) with the auxiliary  $u-$  <CV +u->. Starting from Manichaean (i.e. early) Uyghur times (Clauson 1972: 330), this marker of general possibility has been supplemented by another construction of one of these converbs,  $-(y)V$ , with the auxiliary  $bil-$  ‘to know’:  $-(y)V bil-$ . The latter item gradually broadened its scope with the result that the former almost completely disappeared by the end of the 11<sup>th</sup> century except for some marginal remnants (such as in Western Oghuz, where a reduced form survives as the negation of  $-(y)Abil-$  to the present day). In some varieties of Middle Turkic, a hitherto unattested possibility marker turns up, which consists of the converb in  $-A$  and the auxiliary  $al-$  (usually ‘to take’) and which has developed into the general marker of possibility in a large number of modern Turkic languages, especially in the Kipchak and Chaghatay groups. The earliest attestations of  $-A al-$  are found in the Codex Cumanicus (Kipchak Turkic, early 14<sup>th</sup> c.; cf. Clauson 1972: 124). This marker is also attested in all stages of Chaghatay (15<sup>th</sup>–19<sup>th</sup> c.), where it coexists with  $-A bil-$  (i.e. a later variant of  $-(y)V bil-$ ) in some varieties. The manner in which  $-A bil-$  and  $-A al-$  divide the domain of possibility in these varieties and how the process of replacement evolved is not yet understood.

In the Baburnama, a voluminous Chaghatay text from the early 16<sup>th</sup> c. by an author from Andijan, there is no occurrence of  $-A bil-$  at all. Given the size of this document and the richness in modal expressions found there (see Rentsch (in print)), we may assume that  $-A bil-$  was completely non-existent in the spoken variety underlying this text, implying that  $-A bil-$  (if it had existed in the antecedent to this particular variety at all) must have been replaced very rapidly. On the other hand, Kazakh displays a very striking distribution of  $-A bil-$  and  $-A al-$  to the effect that  $-A bil-$  represents the positive [+internal] member of the opposition [ $\pm$ internal] and  $-A al-$  the negative [–internal] member, a circumstance which seems to run counter to our expectation that the younger form encodes the more marked feature. (The same distribution obviously exists in some other Kipchak languages, such as Tatar (see Thomsen 1959: 420).) However, there is another possible solution to this mystery, namely that  $-A bil-$  is not a surviving old form in Kazakh, but a renewal that incidentally shares the same shape as the old item  $-(y)V bil-$ , which is obsolete in that language or even never existed there. (The converb in  $-A$  is quite productive in Kazakh compared to most other modern Turkic languages, rendering this option possible.) This solution would be in line with the fact that  $-A al-$  as a general marker of possibility (i.e. [–internal]) is actually supplemented by [+internal] renewals involving  $bil-$  in other languages as well. In Uzbek,  $-A al-$  is supplemented by  $-(I)shni bil-$  as a contrasting [+internal] item (see ex. (5)). This construction combines a verbal noun ( $-(I)sh$ ) in the accusative with the auxiliary  $bil-$  ‘to know’ and typifies the same deep structure as Turkish  $-mAyI bil-$ : <VN-ACC+bil->.

- (5) *Men she’r yozishni bilmayman. U narsa q[o’]limdan kelmaydi.*  
 ‘I cannot write poetry. I am not capable of it.’  
 (Web)

$-A bil-$  is said to turn up occasionally in literary Uzbek as well (Kononov 1960: 201), but in reality it is hardly ever encountered in modern Uzbek texts. It might creep into texts by writers with an Oghuz or Kipchak dialectal background.



From a universal perspective, expressions of internal possibility that build on grammaticalized constructions involving verbs of knowing are so widely disseminated (compare French *savoir*, Hungarian *tud*, Mongolian *mede-*, German *können* etc.; cf. also Bybee & Perkins & Pagliuca 1994: 190) that it is not inconceivable that the pattern *-A bil-* has, independently of the Old Turkic renewal in *-(y)V bil-*, been re-inaugurated in a recent renewal of internal possibility in Kazakh and some other Kipchak languages.

### Lexical and idiomatic expressions of modality

Lexemes with modal meaning can function as adverbs, attributes, nouns and predicates in the Turkic languages. Modal adverbials prototypically encode epistemic notions, but occasionally adverbial expressions may also encode event modality, such as the expression *zarūr(at) bolup* 'out of necessity', which occurs several times in the Baburnama. Many epistemic adverbs in Turkic are borrowings, such as Turkish *belki* 'perhaps' (from Persian) and *muhtemelen* 'probably' (from Arabic). These are uninteresting in terms of internal Turkic lexicalization processes. Of greater interest are examples like the Turkish adverb *herhalde*, which is a lexicalized complex comprising a Persian (*her* 'every'), an Arabic (*ḥāl* 'situation') and a Turkic (*-de*, the locative) element. The foreign elements were borrowed into Turkish independently prior to their composition. Originally, *her halde* denoted 'in any case', a meaning that occasionally surfaces even today. More common, however, is the usage of *herhalde* as an epistemic satellite that covers a spectrum roughly ranging from 'certainly' to 'perhaps', i.e. from convinced to undecided (where 'certainly' is the more original epistemic value, derived directly from the literal meaning by metonymy).

Kazakh has a borrowed adjective *mümkin* 'possible' (originally Arabic), which, besides its (attributive and predicative) adjectival uses, can also function as an epistemic sentence adverb, i.e. a modifier of the predicate, hence demonstrating an internal Turkic development (6).

(6) *Mümkin, olar üyde žoq.*

'Perhaps they are not at home.'

(Balakaev & Baskakov & Kenesbaev 1962: 422)

Processes of lexicalization may be reflected not only in semantic or functional changes and/or the fusion of formerly plurimorphemic entities but also in the preservation of an archaic form. In the Anatolian (Western Oghuz) Dede Qorqud stories, the extant redactions of which date from the 16<sup>th</sup> century, we encounter several instances of the lexeme *bolayki(m)*, which is typically constructed with another verb in the optative (7). This lexeme derives from the Turkic stem *bol-* 'to become, to be' in the optative (*\*-GAy*) with the complementizer *ki(m)* added. In 16<sup>th</sup>-century Western Oghuz, *bol-* had developed into *ol-* long before, and the optative was no longer *-(y)Ay* but *-(y)A*. Both *bol-* and *-(y)Ay* survive only in the word form *bolayki(m)*. (The updated form *olaki(m)* occurs with the same function as well.) In terms of meaning, this lexeme, maintaining most of the original semantics of its components ('may it be that'), can reinforce the optative of the verb with which it combines in a sense close to English *hopefully*, but it can also serve as an epistemic satellite (*maybe*). The syntactic behavior of *bolayki(m)* clearly diverges from

what one would expect from its former nature as a matrix clause, as it is to a certain extent mobile within the clause and not confined to the initial position.

- (7) *Gel gēdelüm ikisinden biri bolaykim jān'in vēre, alg'il, menüm jānumi qoğil.*  
 'Maybe one of these two will give his soul. Take it and leave my soul alone!'  
 (Dede Qorqud 83a6–7; Tezcan & Boeschoten 2001: 118)

In some Oghuz varieties this item displays a remarkable perseverance. It is even attested in the Western Rumelian dialect of Vidin (Bulgaria, 20<sup>th</sup> century) with precisely the same function as in the Dede Qorqud stories:

- (8) *Memnun olduq, bolay-ki ağlamayasın qizim.*  
 'We are happy; hopefully, you will not cry, my daughter.'  
 (Németh 1965: 148)

Idiomatic expressions can demonstrate various degrees of conventionalization as well, ranging from a more or less significant increase in frequency to a more or less striking semantic shift or the fossilization of the material structure of items involved in the idiom. All the problems mentioned above regarding the establishment of the degree of conventionalization apply. Within the Islamicized Turkic languages, a construction that entails the adjective 'ājiz 'weak', which is a borrowing from Arabic, and a verbal noun in the ablative displays a remarkable dissemination throughout time and space. It is attested e.g. in Chaghatay (Baburnama, 16<sup>th</sup> c., ex. (9)), Uzbek (10) and Turkish (11). In each of these languages, the morphological components have been adapted to the regulations of the language. That is, the Baburnama and Modern Turkish use the verbal noun in *-mAK*, while Modern Uzbek uses the verbal noun in *-(I)sh* (which has developed from a derivational suffix to an inflectional one in South East Turkic – an example of grammaticalization [context expansion, cf. Himmelmann 2004: 31–34]); Chaghatay employs the ablative in *-din*, Turkish and Uzbek in *-DAñ*; and the phonetic shape complies with the specific requirements of the individual languages as well. It is only the deep structure and the modified semantic design (a shift from *weak* to *unable*) that have been conventionalized. While this expression is not very frequently encountered in any language, it is clearly very persistent.

- (9) *Torgamaqdin 'ajizdurmēn.*  
 'I cannot carve [the goose].'  
 (Baburnama 190b13)
- (10) *Bunga ham sizning bilan bo'lg'an oshnalig'im ko'mak berganlikdan sizga minnatdorlik qilishdan ojizdirman.*  
 'I cannot thank you enough for the fact that my acquaintance with you has helped me in this matter.'  
 (Qodiriy 1926 [1994]: 253)
- (11) *Adamın hekimliğine güvense ciddi olanları da gönderirdi, ama Nurettin Bey basit bir anjine veya bağırsak enfeksiyonuna bile teşhis koymaktan acizdi.*  
 'Had he trusted in that man's medical skills, he would have sent serious cases to him as well, but Nurettin Bey was unable to diagnose a simple angina or an intestinal infection.' (Tunç 2009: 25)



Fossilized deep structures with updated surface structures are not uncommon in the Turkic languages (compare Chaghatay *eligdin kel-*, Turkish *elinden gel-* and Uzbek *qo'lidan kel-* 'to be capable', literally 'to come from one's hand') and form part of a shared idiomatic heritage.

Some idioms can be quite opaque in meaning, such as constructions comprising a future verbal noun with a possessive marker and a form of the verb *kel-* 'to come', which are very widely disseminated throughout Turkic, e.g. Turkish *-(y)AsI gel-* or *-(y)AcAđI gel-*, Chaghatay *-GUsI kel-*, Uzbek and Kazakh *-GIsI kel-*. These constructions denote 'to want', a meaning that is hardly transparent from the meaning of its components.

### Recursion

Lexical and grammatical items with a sufficient degree of conventionalization may be involved in the formation of new idiomatic and grammatical expressions; that is, both lexical and grammatical items can be re-assembled into new (arbitrary) constructions that can undergo either lexicalization or grammaticalization processes. For instance, the Turkish grammatical possibility marker *-(y)Abil-*, which, as we have already seen, derives from a morphosyntactic construction of a grammeme and a lexeme, has combined with the verb stem *ol-* 'to become, to be' (<*bol-*) to form a lexical unit *olabilir* 'may be', which is quite strongly lexicalized, albeit less flexible in usage than the English lexeme *maybe*. *Olabilir* is in turn incorporated into grammaticalized expressions of epistemic modality (*gelmiş olabilir* 'may have come', *geliyor olabilir* 'may be coming' etc.).

The Turkic stem *bol-* 'to become, to be' has developed a certain general affinity to the notion of possibility, which is mirrored in several lexicalized and grammaticalized expressions. Many Turkic languages have conventionalized the third person intraterminal of *bol-* to designate external possibility and permission, e.g. Turkish *olur* and Modern Uyghur *bolidu* 'alright; it is possible'. This is essentially a lexicalization process. In addition, *bol-* is involved in grammatical constructions that also designate external possibility and permission. There is a whole range of options regarding the segment that links main verb and auxiliary. In Old Uyghur, the construction is *-GAlI bol-* with the converb in *-GAlI*, which originally had a final meaning ('in order to'). This construction survives in Modern Uyghur as well. Kazakh and several other Kipchak languages have *-(U)wGA bol-*, where *-(U)w+Ga* is a verbal noun (<*GU*) in the dative. By analogy with the Kipchak pattern, Modern Uyghur has developed a construction *-(V)šKA bol-*, in which the Kipchak verbal noun *-(U)w* is replaced with another verbal noun *-(V)š*. Another strategy frequently encountered since the Middle Turkic period involves the conditional (*-sA bol-*). This construction is found both in Chaghatay and Middle Oghuz and has become an extremely common entity in the modern Turkic languages even beyond the Oghuz and Chaghatay branches. In some languages this originally event modal item has further developed into an epistemic marker. It is highly probable that these grammatical constructions originate in a prior lexicalization process that opened the meaning of *bol-* towards the domain of possibility. In other words, the diachronic development of these markers entail a lexicalization process and a subsequent grammaticalization process.

### Grammaticalization and lexicalization: an integrated model

It has been mentioned that the “box metaphor” is not entirely adequate to illustrate the relationship between lexicon and grammar. Lexicon and grammar operate in concert to form linguistic expressions. Consequently, both lexical and grammatical items are involved in diachronic processes of conventionalization that result in new grammatical or lexical (or idiomatic) items. Often, the affiliation of a linguistic item to either lexicon or grammar is not even clear. There are certainly many unambiguous cases; for example, there will hardly be any doubt about the affiliation of *-sVn* to grammar and of *belki* and *herhalde* to the lexicon. The issue is more problematic with constructions like Turkish *-(y)AsI gel-* and Kazakh *-GIsI kel-*. Are these more adequately classified as grammatical items, or as idioms and consequently falling within the lexicon box? The involvement of the verb *kel-* ‘to come’ suggests some idiomatic background for this construction type; on the other hand, it displays no fewer grammatical traits than *-(y)Abil-* or *-A al-*.

Semantic criteria do not contribute to solving this problem. It is often claimed that grammatical meaning is abstract, while lexical meaning is concrete, and this actually proves to be true in many cases (for instance, the notion “locative”, for which there is a suffix in Turkic, is more abstract than the notion “here”, which is expressed by lexemes like *burada*, and the same is true for the relation between *idi/-(y)di* (“past”) and *dün* ‘yesterday’). The matter is much more complicated in the domain of modality, however. In many cases, the same semantic effect can be achieved by either grammatical or lexical (idiomatic) means, e.g. *belki* vs. ASP+*olabilir* ‘maybe’, *muhtemelen* vs. *-mAsI muhtemel* ‘probably’ and *-mAktAn aciz* vs. *-mAyI bilme-* (internal possibility). Distinguishing grammar and lexicon (and, by extension, grammaticalization and lexicalization) only from a semantic point of view is hardly possible. Frequency is also not a reliable criterion. Grammaticalization is claimed to involve an increase in frequency, but lexical items can be either frequent or infrequent; that is to say lexicalization does not necessarily entail an increase in frequency (Himmelmann 2004: 37). Frequency does not contribute much to the picture apart from the fact that conventionalized items are usually more frequent than arbitrary ones (except conventionalized items on the decline, i.e. dying forms).

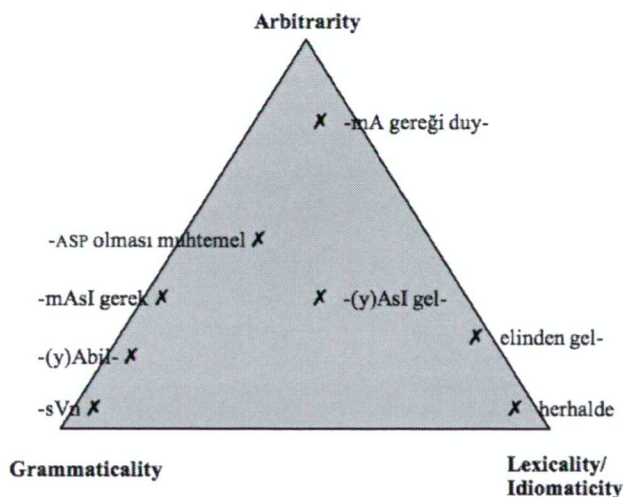
Rather than conceptualizing grammar and lexicon as completely separate categories (“two boxes”), it is more adequate to imagine them as a continuum (“one big box”), with relatively clear cases towards both ends of the spectrum (such as inflectional morphology at the one end and free items at the other) and more ambiguous cases in between. Lexicalization and grammaticalization would then enter a complex relationship both mutually and with arbitrariness as their common starting point, in which both processes may operate either in concert or subsequently and yield both very different and very similar results. In some cases, it will be impossible to distinguish both processes. The relationship between the three poles arbitrariness, grammaticality and lexicality/idiomaticity, according to this view, can be visualized as a triangle (Figure 1). The representation as a triangle accounts for the fact that grammaticalization and lexicalization share a common starting point. The less advanced the process of conventionalization is, the more obscure the direction the process will take. Even at a later point, a conventionalized item may hover somewhere between grammar and lexicon. Needless to say, both types of conventional-



ization may stop before reaching the base of the triangle, irrespective of whether the item survives or disappears. Conventionalized items are again “arbitrary” in that they can productively combine with other morphemes and/or lexemes to form utterances. Recursion has to be imagined as a return to the starting point. That is, while the third person volun-tative in *-sVn* has to be situated near the base of the triangle, a construction like *-sVn iste-* ‘to want somebody to X’ (12) has to be considered the result of a new conventionalization process with arbitrariness as the starting point.

- (12) *Sana bu yolda sonuna kadar yoldaş olayım isterim ama yapamam.*  
 ‘I would love to accompany you on this path until the end, but I cannot do it.’  
 (Şafak 2009: 257)

Figure 1. Conventionalization triangle with selected Turkish items



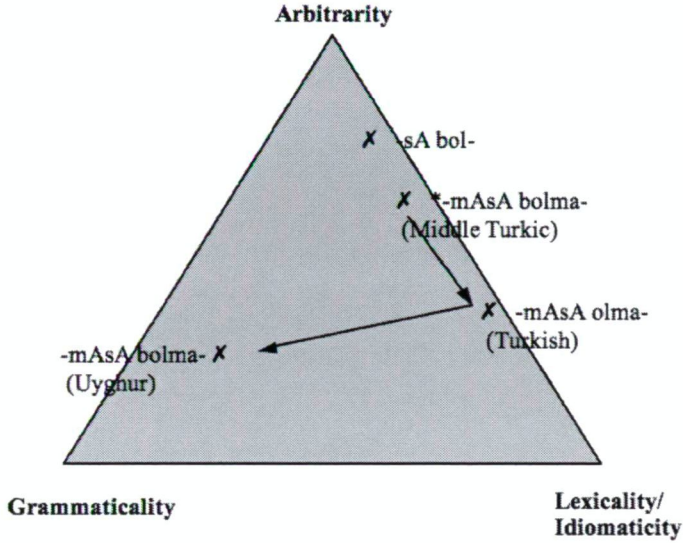
Grammaticalization and lexicalization are usually considered irreversible (principle of unidirectionality). However, they need not evolve rectilinearly. A given construction may first develop towards idiomatity and be integrated into a grammatical paradigm at a later point. This seems to have happened with the Turkic construction *-mAsA bolma-* <NEG-COND+be-NEG>, which is based on the expression of possibility *-sA bol-* which has already been mentioned. *-mAsA bolma-* displays a wide dissemination throughout the Turkic world, but varies in its exact profile. In Turkish (13), it is a relatively weakly conventionalized construction that occurs quite infrequently and is semantically very close to the meaning of its components. Example (13) could be rendered as ‘If I don’t do it, it is impossible’, which is almost a literal translation, or slightly modified as ‘I cannot help but do it.’ In several Central Asian Turkic languages, especially Kazakh, Kirghiz, Uzbek and

Uyghur (14), this construction has developed further to designate strong necessity ('to have to do something by all means'; cf. Yakup 2009: 487–489). With this semantic profile it supplements less specific markers of general necessity consisting of a verbal noun with a possessive marker and \**kergek* <VN-POSS+kergek>.

- (13) *Yapmasam olmaz.*  
 'I cannot help but do it.'  
 (Ümit 2003: 209)
- (14) *Bügün buni öginip bolmisam bolmaydu.*  
 'I have to memorize this today by all means.'  
 (Abduveli 2001: 2)

The Turkish and the Central Asian variants, despite representing exactly the same morphological structure, differ greatly in their degree of conventionalization and in their location between lexicon and grammar. Although frequency has been discarded as a sufficient criterion for establishing the degree of conventionalization, it is notable that the construction is significantly more frequent in Central Asian Turkic than in Turkish (roughly by a factor of 100 in Uyghur and Kazakh), a dimension that strongly speaks in favor of a higher degree of conventionalization in CAT. The increase in frequency is likely to be a result of language contact, as <NEG.COND+be-NEG> constructions are strikingly common in the area (attested e.g. in Manchu and several Mongolic languages; the underlying logic pattern < $\neg\Diamond\neg p$ > (cf. de Haan 2006: 55) is attested in Classical Chinese since at least the 5<sup>th</sup> century BC). Moreover, the CAT variant has been integrated into a paradigm of MOD-1 operators that enter into specific semantic oppositions (such as necessity vs. possibility, strong vs. weak etc.), indicating a stronger degree of grammaticalization than for the Turkish variant. Figure 2 illustrates the path of development for the Turkish and the Uyghur variants of *-mAsA bolma-*. The emergence of this structure presupposes a conventionalized pattern *-sA bol-* 'to be possible' < $\Diamond p$ >, to which negation suffixes could be added in a productive way, rendering *-sA bolma-* 'to be impossible' < $\neg\Diamond p$ >, *-mAsA bol-* < $\neg\neg p$ > and *-mAsA bolma-* < $\neg\Diamond\neg p$ >. Each of these structures was eligible for conventionalization, and in fact *-sA bol-* has been conventionalized in very many Turkic languages to different extents (the path for this item is not indicated in the figure). The Turkish expression *-mAsA olma-* is certainly no longer arbitrary, but it is neither particularly widely used, nor has it been integrated into the grammatical system of the language. The Uyghur variant is both much more prolific and enters systematic oppositions to other grammatical markers. Hence, its conventionalization is not only more advanced but has also developed in a direction different from the Turkish counterpart.



Figure 2. Conventionalization path of *-mAsA bolma-*

### Acknowledgement

The research for this paper was made possible by a grant from the Alexander von Humboldt Foundation.

### Abbreviations

ACC	Accusative	MOD	Modality
ASP	Aspect	N	represents a noun stem
CAT	Central Asian Turkic	NEG	Negation
COND	Conditional	POSS	Possessive
CV	Converb	VN	Verbal noun
DAT	Dative	X	represents a verb stem

### Sources

- Abduveli, Hüsenjan 2001. *Közge aylanyan köñül*. Ürümçi: Şinjan xelq neşriyati.  
 Baburnama: Beveridge, A. S. (ed.) 1905. *The Bâbar-nâma. Fac-simile*. Leyden & London.  
 Qodiriy, Abdulla 1926/1929 [1994]. *O'tkan Kunlar/Mehrobdan Chayon*. Toshkent: G'afur G'ulom nomidagi adabiyot va san'at nashriyoti.  
 Şafak, Elif 2009. *Aşk*. İstanbul: Doğan Kitap.  
 Tunç, Ayfer 2009. *Bir deliler evinin yalan yanlış anlatılan kısa tarihi*. İstanbul: Can Yayınları.

Ümit, Ahmet 2003. *Beyoğlu Rapsodisi*. İstanbul: Doğan Kitap.

Ümit, Ahmet 2006. *Kavim*. İstanbul: Doğan Kitap.

Ümit, Ahmet 2008. *Bab-ı Esrar*. İstanbul: Doğan Kitap.

Web: <http://forum.oqayiq.uz/viewtopic.php?t=326&sid=309c486b4d6c3dc737b49ed6e10c73f4>  
(retrieved on January 12, 2011).

## References

- Balakaev, M. B. & Baskakov, N. A. & Kenesbaev, S. K. (eds.) 1962. *Sovremennyy kazaxskij jazyk. Fonetika i morfologija*. Alma-Ata: Izdatel'stvo Akademii Nauk Kazaxskoj SSR.
- Bybee, J. & Perkins, R. & Pagliuca, W. (eds.) 1994. *The evolution of grammar. Tense, aspect and modality in the languages of the world*. Chicago & London: The University of Chicago Press.
- Clauson, G. 1972. *An etymological dictionary of pre-thirteenth-century Turkish*. Oxford: Clarendon.
- de Haan, F. 2006. Typological approaches to modality. In: Frawley, W. (ed.). *The expression of modality*. Berlin & New York: Mouton de Gruyter. 27–69.
- Himmelman, N. P. 2004. Lexicalization and grammaticization: Opposite or orthogonal? In: Bisang, W. & Himmelman, N. P. & Wiemer, B. (eds.). *What makes grammaticalization? A look from its fringes and components*. Berlin & New York: Mouton de Gruyter. 21–42.
- Kononov, A. N. 1960. *Grammatika sovremennogo uzbekskogo literaturnogo jazyka*. Moskva: Izdatel'stvo Akademii Nauk.
- Lyons, J. 1977. *Semantics*. Vol. 2. Cambridge: Cambridge University Press.
- Németh, J. 1965. *Die Türken von Vidin. Sprache, Folklore, Religion*. Budapest: Akadémiai Kiadó.
- Nuyts, J. 2006. Modality: Overview and linguistic issues. In: Frawley, William (ed.). *The expression of modality*. Berlin & New York: Mouton de Gruyter. 1–26.
- Palmer, F. R. 1986 (2001). *Mood and modality*. Cambridge: Cambridge University Press.
- Rentzsch, J. 2010. Zur Modalität im Türkischen. In: Boeschoten, H. & Rentzsch, J. (eds.). *Turcology in Mainz/Turkologie in Mainz*. Wiesbaden: Harrassowitz. 209–224.
- Rentzsch, J. (in print). Modality in the Baburnama. *Turkic languages*.
- Tezcan, S. & Boeschoten, H. 2001. *Dede Korkut Oğuznameleri*. İstanbul: Yapı Kredi Yayınları.
- Thomsen, K. 1959. Das Kasantatarische und die westsibirischen Dialekte. In: Deny, J. et al. (eds.). *Philologiae Turcicae fundamenta 1*. Aquis Mattiacis. 407–421.
- van der Auwera, J. & Plungian, V. A. 1998. Modality's semantic map. *Linguistic Typology* 2, 79–124.
- Yakup, A. 2009. Necessity operators in Turkish and Uyghur. *International Journal of Central Asian Studies* 13, 479–493.