Layered Structure and Reference in a Functional Perspective

edited by Michael Fortescue
Peter Harder
Lars Kristoffersen

Offprint

This is an offprint from:

Michael Fortescue, Peter Harder and Lars Kristoffersen
LAYERED STRUCTURE AND REFERENCE
IN A FUNCTIONAL PERSPECTIVE
John Benjamins Publishing Co.
Amsterdam/Philadelphia
1992
(Published as Vol. 23 of the
PRAGMATICS & BEYOND NEW SERIES)
ISSN 0922-842X
ISBN 90 272 5035 9 (Eur.) / ISBN 1-55619-291-6 (U.S.)

The Treatment of Turkish Nominal Compounds in FG

Gerjan van Schaaik
Dept. of Computational Linguistics
University of Amsterdam

0. Introduction

The aim of this paper is to investigate how Nominal Compounds of Turkish can be treated within the framework of Functional Grammar, as developed by Dik (1978, 1989) and others. The paper is organized as follows. In section 1 the relevant constructions concerned will be introduced, together with some terminological distinctions that are relevant for a further exploration. It will be shown that a first approximation towards a generative model is inadequate, because the rule proposed is too powerful. Section 2 introduces some linguistic data that will determine the direction in which a possible solution must be sought. In section 3, this model will be further elaborated and evaluated in the light of complex compounds. In section 4, I will discuss the relation between the processes of nominal compounding and nominal derivation, and in section 5 I will go into items related to the lexicalization of compounds. Finally, section 6 summarizes the results obtained.

1. Nominal compounds

Turkish nominal compounds generally consist of two or more nouns, as can be exemplified by the following:

(1) a. *çay ev-i*
tea house-CM
'tea house'

b. *türk dil-i gramer-i*
turk language-CM grammar-CM
'Turkish language grammar'

c. *türk dil kurum-u*
turk language society-CM
'Turkish language society'

Nominal compounds as (1a) will be referred to as Simple Compounds, whereas the types represented in (1b) and (1c) are Complex Compounds since they contain more than two nouns. Compounds may contain a suffix that is usually glossed as the 'possessive suffix third person singular', but, as will be shown in section 2, this term is in most cases not very appropriate due to the fact that a possessor is usually not present. Hence, the suffix will be referred to as the Compound Marker (CM).

The constructions under consideration are those which are traditionally known as the 'indefinite izafet constructions' (Lewis 1967; Gadžieva 1973; and Baskakov 1975), as opposed to the so-called 'definite izafet' constructions. The latter type can be represented by (2b), and its 'indefinite' counterpart as (2a):

(2) a. *ev sahib-i*
house-owner-CM
'(a/the) house owner'

b. *ev-in sahib-i*
house-GEN owner-POSS3s
'the owner of the house'

On the basis of pairs as in (2), it has always been thought that there is some semantic or syntactic relation between both construction types. It is quite easy to understand why: the suffix *-i* in (2a) and (2b) is identified as one and the same possessive marker. Therefore, both constructions differ only in one respect, namely, *ev-in* 'house-GEN' (2b) contains a genitive suffix whereas *ev* in (2a) does not. *Sahib* 'owner' of (2b), then, is modified by a particular *ev* (indicated by the genitive) and the whole is interpreted as 'definite'. Since there is no such marker in construction (2a) it must be 'indefinite'.

However, there are quite a number of arguments in favour of an alternative view.

First of all, there is no semantic relation whatsoever between (2a) and (2b). Indeed, (2a) and (2b) are built up by the same lexical material, but that is sheer coincidence. In (2a) *sahib* 'owner' is modified by *ev* 'house' in such a way that the set of possible referents of *sahib* is restricted to 'house owners', thereby excluding all other types of owners. Construction (2b), on the other hand, represents a pure possessive construction: in terms of FG, *sahib* 'owner' is restricted by a possessor term. This term has been assigned the semantic function POSS[1], which is expressed as a genitive case marker.

Secondly, pairs like (2a) and (2b) can be formed only with a very limited number of nouns. Taking for instance the lexical material of (1a) and making a construct analogous to (2b), we get something quite ungrammatical: *çay-ın ev-i*. Hence, such oppositions are to be considered as pseudooppositions and they should not play (too) a great role in grammatical analyses.

A third reason for a sharp distinction between (2a) and (2b) is that compounds (2a) can only be modified as a whole, whereas both members of (2b) can be modified by adjectives, numerals, and participles.

(3) a. **misafir küçük oda-sı*

küçük misafir oda-sı
small guest room-CM
'a/the small guest room'

b. **misafir bir oda-sı*

iki misafir oda-sı
two guest room-CM
'two guest rooms'

c. **misafir kal-acağ-ınız oda-sı*

kal-acağ-ınız misafir oda-sı
stay-PartFut-'you' guest room-CM
'the guest room where you will stay'

This observation justifies the conclusion that we are dealing with nouns that constitute firmly connected units. As a first abstraction from the data, Turkish compounds can be represented as follows:

(4) $noun_1$_$noun_2$- CM

The underscore in $noun_1$_$noun_2$-CM indicates that both nouns form a tightly-knit unit, since other lexical elements cannot occur in the position between the two nouns. Now, if we want to relate (4) to an operational generative model, we could assume that it is the output of a Predicate Formation Rule that accepts $noun_1$ and $noun_2$ as input. Indeed, this mechanism has proven to be able to give an adequate description in a great number of cases in a variety of languages (cf Dik 1989, 1980:25f, 90f; De Groot 1987; MacKenzie 1987; Van Schaaik 1985). For Turkish Nominal Compounds the general structure of such a rule would be of the following shape:

(5) Compound Formation Rule

Input:	$noun_1$ (x_i) $noun_2$ (x_j)	e.g.	*misafir*$_N$ (x_i) *oda*$_N$ (x_j)
Output:	$noun_1$_$noun_2$-CM (x_k)		*misafir_oda*$_N$*-sı* (x_k)

This rule works as follows. There are two input predicates: a $noun_1$ (x_i) and a $noun_2$ (x_j) that are output as the compound $noun_1$_$noun_2$-CM (x_k). This output predicate has only one argument position and the strings $noun_1$ and $noun_2$ are concatenated (indicated by '_'). These characteristics are formally expressed by the Compound Marker (CM). The form of the CM is determined by the vowel quality and the structure of the final syllable of $noun_2$: if $noun_2$ ends in a vowel the CM becomes *-sI*, otherwise *-I* occurs. The *I* is expressed as 'i' following a front_flat vowel (*e* or *i*), as 'ü' after a front_round vowel (*ö* or *ü*), as 'ı' after a back_flat vowel (*a* or *ı*), and as 'u' after a back_round vowel (*o* or *u*). As a matter of fact, the rule as outlined above it too powerful. In order to support this statement, we will introduce some linguistic data from three different domains.

First, in a construction based on a compound the CM does not always occur. When a compound term is (underlyingly) restricted by a possessor term, possessive concord must be expressed by way of a possessive marker. That is, a marker that

agrees in grammatical person with that particular possessive restrictor. This marker excludes the CM. The entire paradigm for *diş_fırça* 'tooth brush' and *araba* 'car' are given below:

(6) a. *(ben-im)* *diş_fırça-m* b. *(ben-im)* *araba-m*
I-GEN tooth brush-POSS1s I-GEN car-POSS1s
'my tooth brush' 'my car'

(sen-in)	*diş_fırça-n*	*(sen-in)*	*araba-n*
(on-un)	*diş_fırça-sı*	*(on-un)*	*araba-sı*
(biz-im)	*diş_fırça-mız*	*(biz-im)*	*araba-mız*
(siz-in)	*diş_fırça-nız*	*(siz-in)*	*araba-nız*
(onlar-ın)	*diş_fırça-ları*	*(onlar-ın)*	*araba-ları*

Thus, when a possessive term is based on a noun, the noun (Simplex or Compound) has a possessive marker that agrees with the grammatical person of the possessor, but when a Compound Noun is used as a 'free' nominal (cf. 1a), a suffix (*-sı/-ı*) is attached that is morphologically identical with the possessive suffix third person singular. However, from a semantic point of view, this should be regarded as a coincidence, because there is no reason to assume that a possessor is involved in the process of compounding. Therefore, I think it is useful to distinguish between possessive markers and the compound marker, also because they mutually exclude one another:

(7) a. **diş_fırça-sı-m*
tooth brush-CM-POSS1s
'my tooth brush'

b. **diş_fırça-sı-sı*
tooth brush-CM-POSS3s
'her/his tooth brush'

c. **Bulgaristan-ın (İstanbul Başkonsolosluğ-Ø)-u*
Bulgaria-GEN (Istanbul Consulate-General-Ø)-POSS3s
'The Istanbul Consulate-General of Bulgaria'

Of course, a construction like *diş_fırça-sı* is ambiguous, due to the fact that the (pronominal) possessor can be unexpressed. That is why we get two (contextually determined) interpretations: 'tooth brush' and 'her/his tooth brush'. Example (7c) also clearly shows the mutual exclusion of the CM and the possessive suffix (POSS3s): the 'possessor' (in casu *Bulgaristan* 'Bulgaria') is expressed in the genitive (GEN), and, because of 'possessive concord', the 'possessed' obtains the suffix POSS3s. The CM is not present (as indicated by -Ø).

The general objection to be raised against (4) can be formulated as follows: In (4) the CM is generated even if it cannot be expressed. One solution is to delete it, but that is an operation which is not allowed within the framework of Functional Grammar.

The second objection against (4) concerns another kind of operation which FG rejects. This is the case of a plural expression of (1a):

(8) *çay_ev-ler-i* Compare: *çay_ev-i* (= 1a)
tea house-plur-CM

One might claim that the plural marker could easily be inserted, but on the basis of the next example it will be clear that more operations will be involved:

(9) *at_araba-lar-ı* Compare: *at_araba-sı*
horse cart-plur-CM

If *at_araba-sı* 'horse cart-CM' were to be the basic form for the development of (9), at least the CM *-sı* would have to be removed and, in one way or an other, be adapted to a new (morpho)phonological environment. In that case, it would have to be adjusted to the required form *-ı*. That, again, would imply that a kind of transformation must be carried out.

Finally, similar operations are necessary as long as we take a compound including the CM as a starting point in the process of nominal derivation. It is I think sufficient here to give a few examples:

(10) a. *koyun_baş-lı* Compare: *koyun_baş-ı*
sheep-head-Adj sheep-head-CM
'mutton headed' 'sheep head'

b. *güneş_göz-lük-lü* Compare: *güneş_göz-lüğ-ü*
sun glasses-Adj sun glasses-CM
'with sun glasses' 'sun glasses'

With respect to the output of a Predicate Formation Rule, there is only one way out. On the basis of the foregoing, we have to assume that such a rule produces a compound predicate *without* a Compound Marker. Doing so would imply that the grammar (the relevant part of the expression rules) would need an extra condition that is tested before a compound predicate can be expressed. This condition can be formalized as follows: if a compound predicate is to be expressed as a term, the CM must be attached. Otherwise, test whether it is restricted by a possessor term and in that case a possessive marker will follow. Thus, (4) can be rewritten as:

(11) $noun_1$_$noun_2$-

2 Terms as Input?

The objections raised so far can be countered by the assumption that the output predicate of the proposed Predicate Formation Rule is a bare nominal predicate. With respect to the input of the rule, however, there are some data not yet introduced which strongly suggest a further adaptation of (11). In its present form (that is, taking the requirement discussed in section 1 into account), it will accept two nominal predicates as its input. For a number of cases the rule will indeed produce correct output, as can be exemplified by (1a) and (1c). The construction of (1b), however, cannot be accounted for. We will show them again, followed by a short discussion.

(12) a. *çay ev-i* (=1a)
tea house-CM

b. *türk dil-i gramer-i* (=1b)
Turk language-CM grammar-CM

c. *türk dil kurum-u* (=1c)
Turk language society-CM

Example (12a) represents a Simple Compound: *çay* 'tea' and *ev* 'house' are combined to a new notion *çay_ev-* 'tea house', which is expressed as a term, to be interpreted as 'the/a tea house'.

A compounding rule is cyclic: in principle it can be applied an infinite number of times (on its own output). In (12c), for instance, the rule has been applied twice. First of all, *dil* 'language' and *kurum* 'society' are fused into *dil_kurum-* 'language society'. This initial product is subsequently re-inputted as $noun_2$ in the rule together with *türk* 'Turk' as $noun_1$, giving (12c). Finally, the compound has been expressed as a term, marked by the CM. The cyclic nature of the rule can be represented in the (final) output by putting all intermediary results between brackets. Hence, we get for (12c):

(13)

$$\left[\left(\textit{türk\_(\ dil\_kurum\ )}_{Nc} \right)_{Nc} \textit{-u} \right]_{T}$$

Now that we have seen that the CM occurs only when a compound is expressed as a term, it might be worth while examining if this could give a clue to the solution of the problem with (12b). Since it contains a CM on *dil* 'language', the question arises whether *türk dil-i* 'Turkish language' of (12b) should be considered a term rather than a noun. If this is actually the case, it implies that at the first stage of compounding, *türk* 'Turk' and *dil* 'language' together constitute the predicate *türk_dil-* 'Turkish language', being expressed as the term (*türk_dil-i*). In the second application of the rule, this term has been re-inputted, together with *gramer* 'grammar', eventually resulting in the term *türk dil-i gramer-i* (=12b) via the predicate *türk_dil-i_gramer-*. Its structure can be depicted by:

(14)

$$\left[\left(\left[\ (\ \textit{türk\_dil}\ )_{Nc} \textit{-i} \right]_{T} \textit{\_gramer} \right)_{Nc} \textit{-i} \right]_{T}$$

Before we go into any theoretical considerations concerning the possibility of accepting fully specified terms on the input side of a Predicate Formation Rule, consider the following data:

(15) a. (*kör-ler*) *okul-u*
blind-plur school-CM
'school for the blind'

b. (*üç çocuk*) *anne-si*
3 child mother-CM
'a mother of three children'

c. (*önemli iş-ler*) *dosya-sı*
important matter-plur file-CM
'file of important matters'

d. (*Türkiye-nin ses-i*) *radyo-su*
Turkey-GEN voice-Poss3s radio-CM
'radio "the voice of Turkey"'

e. (*Ankara*) *belediye-si*
Ankara municipality-CM
'the municipality of Ankara'

In (15a) the occurrence of a plural marker (*-ler*) presupposes pluralization of the noun *kör* 'blind' before it is combined with *okul* 'school'. This implies that 'somewhere' in the formation rule a modification must be made that allows for (optional) pluralization of input predicates that will be outputted as the left-most member of the compound. Note that this requirement does not hold for the second input predicate, because a plural form of the entire construction is accounted for on the level of term expression. Clear examples were presented in (8) and (9): (*çay_ev*)-*ler-i* '(tea house)-plur-CM' and (*at_araba*)-*lar-ı* '(horse_cart)-plur-CM'.

However, inclusion of a pluralization operation is not the only necessary extension of the rule. If we look at (15b), it is evident that except for the opposition between singular and plural, we will have to specify an additional rule that deals with everything between 'singular' and 'plural'. One might argue that (15b) is a more or less petrified expression, such as the German (*Drei-groschen*)-*oper*, but speaking in a strict grammatical sense, instead of *üç* 'three' any number could occur in combination with *çocuk* 'child' and *anne* 'mother'. And again, a modification of the

proposed rule would have to be carried out to accommodate this type of quantification.

With examples like (15c) and (15e), it becomes clear that we will probably have to include a rather lengthy series of adaptations to the rule. Apart from the expression of term operators (e.g. the category Number, as in (15a-b)), several syntactic rules must be included as well. In (15c) the adjective *önemli* 'important' is a restrictor on *iş* 'matter'. The whole thing is expressed in the plural and *as such* constitutes the left hand member of the entire compound. Thus we at least need an ordering rule for the placement of an adjectival restrictor. Obviously, a possessive phrase has to be input to the rule in order to produce (15d). Here, too, we need syntactic rules for the expression of such phrases.

The data presented here strongly support the view that not only bare nominal predicates but also fully specified terms can be taken as the input of a Predicate Formation Rule for Turkish compounds. Constructions such as (12b), where the occurrence of a CM in the middle of a compound is the problem, favour this approach, as does the fact that all the extra modifications indicated above are strictly speaking rendered superfluously.

Indeed, it is not necessary to incorporate pluralization, the expression of cardinal numbers, syntactic rules for the placement of an adjective or of a possessor term, since all these can be handled by the expression rules for terms. As a matter of fact, these operations are typical operations in the domain of term expression, which is already an essential part of the grammar. It follows, then, that in the present view both term formation and term expression are a precondition for the formation rules of compounds.

The idea that terms can function as the input of a predicate formation rule is further reinforced by the assumption within the framework of FG (cf. Dik, 1989: 55) that the lexicon contains, apart from predicates, basic terms in the form of pre-modeled term structures.

Clear examples are personal pronouns and proper nouns. Most of the proper nouns refer to definite singular entities, although some of them may have a plural form. For instance, the proper name *Ankara* (cf (15e)) refers to a unique (hence its definiteness) singular entity, such that its plural form would hardly ever be expected.

However, the proper name *the Alps*, has a plural form although it refers to a region that is regarded as a whole. Therefore it is not very likely to ever hear someone refer to an Alp (as a singular name for some mountain located in *the Alps*).

Nevertheless, singularity or plurality of the surface forms is coded in their underlying (basic) term structure, as can be expressed by:

(16) def sing $ankara_{Np}$ → *Ankara*
def plur alp_{Np} → *the Alps*

Similarly, for the proper names of Turkish we may assume that they are stored in the lexicon as basic term structures. Whenever such a term must be expressed, the usual set of expression rules for terms will be applied. Returning to the proposal as given in (11), it is quite easy now to imagine how compounds of Turkish are formed. We use the schema of (5):

(17) Compound Formation Rule

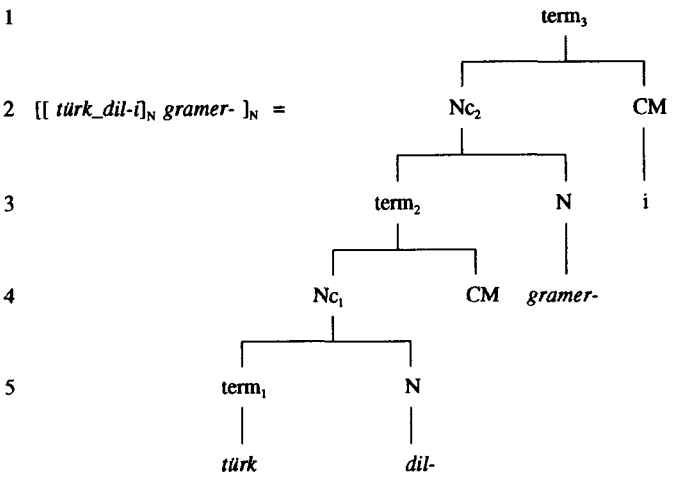
Input:	term	e.g.	def sing x_i : *Ankara* (x_i)
	$noun_2$ (x_j)		$belediye_N$ (x_j)
Output:	$term\_noun_2$- (x_k)		$Ankara\_belediye_N$- (x_k)

3. Evaluation

As in most languages that have the device of compounding as a means of forming new predicates, compounding in Turkish can be done recursively. In principle, compound strings of an infinite length could be produced by the grammar, yet the maximal length is limited for practical (that is: psychological) reasons. Generally speaking, it is hard to indicate how many times a Compound Formation Rule can be applied. Since the Compound Formation Rule, as proposed in (17), has two input items (namely a term plus a nominal predicate), it is clear that two types of (recursively generated) structures are to be distinguished. In this section we will study the so-called Left Branching Compounds and Right Branching Compounds. The construction of (18) is an example of a (complex) compound that is formed on the basis of a term the head of which is itself a compound plus a bare nominal predicate. The structure of the entire compound predicate of (18) can be represented by the tree diagram of (19). Consider the following, a Left Branching Compound:

(18) (*türk dil-i -gramer-*)*i*
turk language-CM grammar-CM
'Turkish language grammar'

(19)



"türk dil-i gramer-i"

At the top level of this tree, the term *türk dil-i gramer-i* is represented by $term_3$. This node has two branches: the right hand side node depicts the compound marker (generated at the top level), and the left branch constitutes the complex compound Nc_4: *türk dil-i gramer-*. As can further be inferred from the structure of (19), this compound was formed by the input ($term_2$) *türk_dil-i* and the nominal predicate *gramer*. The string *türk_dil-i* contains the CM *-i*, because it was input as a term. Finally, *türk_dil-* was brought about by compounding (via Nc_1) the term *türk* and the predicate *dil*. As for the directionality of the formations sketched here, it is a typical bottom-up process: terms and compounds are formed in an alternating order, starting with the formation of $term_1$ which is followed by the formation of the first compound by Nc_1. This sequence is repeated once more, until eventually the top-most term (=$term_3$) is formed.

Another example of the Left Branching type shows clearly that term formation is fundamentally involved in the process of compounding.

(20) *(Sovyet (Sosyalist Cumhuriyet-))-ler-i Birliğ-i*
soviet socialist republic-plur-CM union-CM
'the Union of Soviet Socialist Republics'

$$\Big[\Big(\Big[\Big(\,[sovyet]_T([sosyalist]_T\{cumhuriyet\}_N)_{Nc}\Big)_{Nc}\text{-}ler\text{-}i\Big]_T\{birliQ\}_N\Big)_{Nc}\text{-}i\Big]_T$$

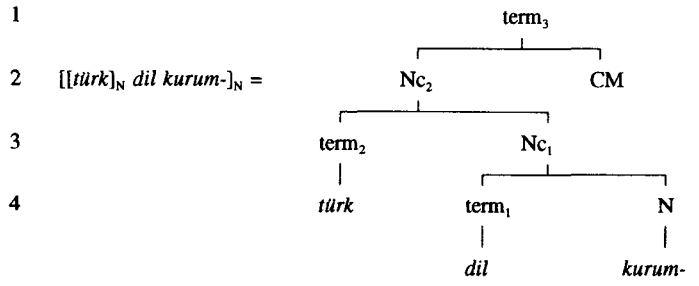
The term *sovyet* is combined with *(sosyalist_cumhuriyet)-* (in itself a compound) into the predicate *(sovyet_sosyalist_cumhuriyet)-*. Next, a term is formed from it which is expressed in its plural form.

Therefore, the plural suffix denotes the plurality of the entire compound *(sovyet_sosyalist_cumhuriyet)-*. Because it is expressed as a term, the CM *-i* follows finally. Compounds of the Right Branching type are, as a matter of fact, built up in a way similar to those of the Left Branching type. The fact that compounds of the Right Branching type differ, however, in both their lexical and surface forms from those of the Left Branching type justifies a separate treatment. The main difference from the Left Branching type is the notable absence of a Compound Marker on any (but the final) constituent. In its simplest form, the Right Branching type can be illustrated by the following example:

(21) *türk (dil kurum)-u*
Turk language society-CM
'Turkish language society'

Again, the syntax of such a construction and the relative dependency of the constituents can be depicted by a tree structure:

(22)



"türk dil kurum-u"

At level 4, a term (*dil*) plus a bare nominal predicate (*kurum*) form the input of the Compound Formation Rule, which outputs the compound *dil_kurum-*. This compound is again inputted to the rule (level 4) together with the term *türk*, which yields the second compound, Nc_2: *türk_(dil_kurum-)*. When this new predicate is expressed as a term (on level 1), the Compound Marker *-u* must appear. The number of formation levels of this type of compounding is always less than that of the Left Branching type, because the nodes indicating pure compounding (levels 2 and 3) are directly connected (Nc_1 - Nc_2), whereas for the Left Branching type of compounding term formation there is the characteristic intermediary stage (Nc_2 - term - Nc_1; cf levels 2, 3, 4 in (19)). Due to the recursive nature of the Right Branching type of compounding, this process can be applied several times to the output of a previous run. The following example demonstrates the recursive power of the rule:

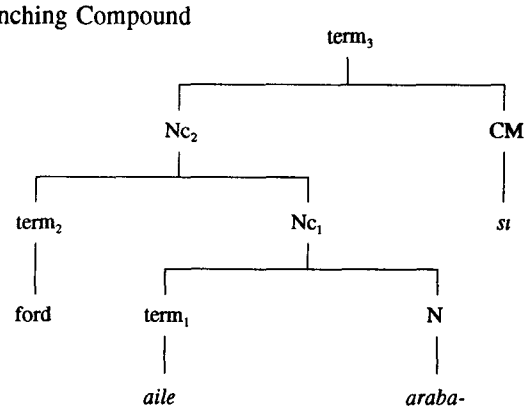
(23) *(İstanbul (orman (bölge (müdürlüğ-)))ü*
Istanbul forest area directorate-CM
'Istanbul regional directorate of forestry'

From (23) it follows again that the Compound Marker can occur only once, due to the fact that a multi-level compound obtains its CM exclusively when a term is based on it. Lewis (1967: 45) states that "It is most important to note that the third person suffix is not repeated though *theoretically* (italics are mine) one might have expected: Ankara (Kız Lise-si)-si " ('Ankara Girls Lycee'). On the basis of the foregoing, however, I think that structures like:

(24) *(Türkiye (cumhuriyet (merkez banka-sı)-sı)-sı)*
Turkey republic center bank-CM
'Central bank of the republic of Turkey'

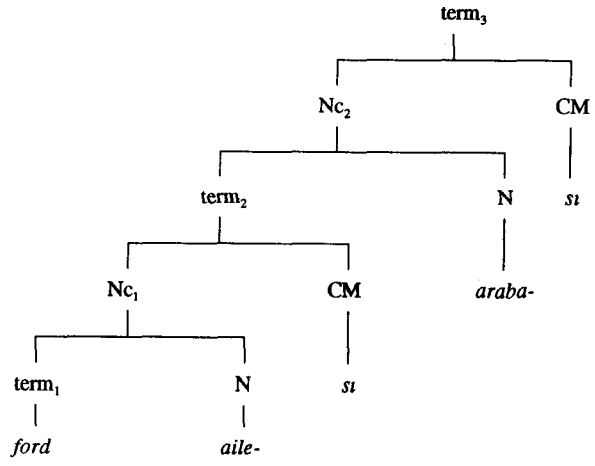
for which one or more deletion rules must be set up in order to avoid several instances of *-sı*, give the wrong idea of the generative history of compounds. The fundamental difference between Left Branching and Right Branching compound can be summarized as follows. Whereas the structure of a Left Branching compound starts with a term and a bare nominal predicate for the formation of the 'deepest' (sub-)compound, the Right Branching type starts at its lowest level with a term and a nominal predicate that is a compound itself. Two examples from Lewis (1967:46) clearly show this difference:

(25) Right Branching Compound



"*Ford aile araba-si*"
'the Ford family-car'

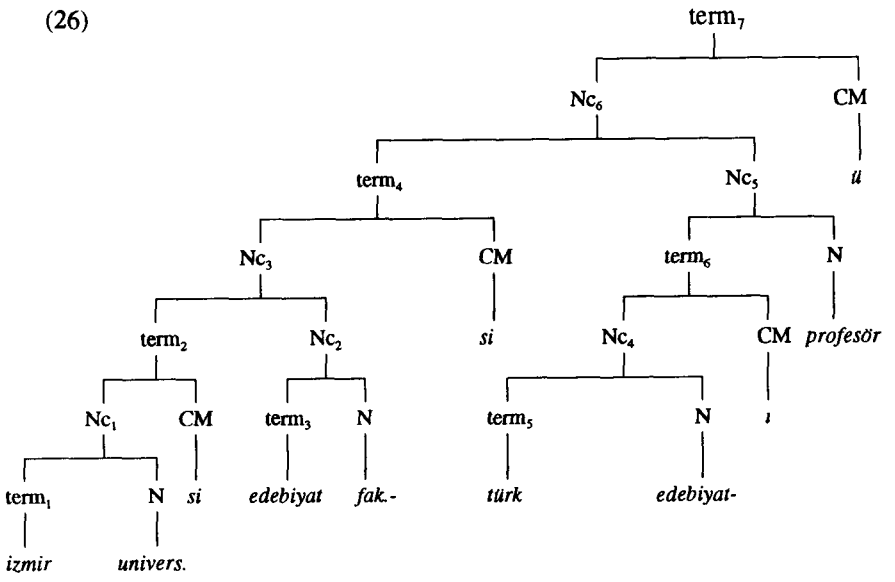
Left Branching Compound



"*Ford aile-si araba-si*"
'the Ford-family car'

The last example we give here shows clearly that the Right Branching type can very well be an element of a Left Branching structure.

(26)



"İzmir Üniversite-si Edebiyat Fakülte-si Türk Edebiyat-ı profesör-ü"
'professor in Turkish literature at the I. University's faculty of arts'

Concluding this section, we may say that there is ample evidence for how the formation of Turkish compounds can best be accounted for. It can be described by a Compound Formation Rule that takes a term plus a nominal predicate as its input, thereby producing a compound nominal predicate. The derived character of this predicate may be signalled by a Compound Marker. This CM is generated at term level, and is expressed only when in the predicate in a term structure is not restricted by a possessor term. Needless to say, the superfluous generation of the CM and its deletion is prevented in the system outlined here.

4. Compounding and Derivation

Additional evidence for a 'late' generation of the CM is obtained from the examination of the relation between nominal derivation and compounding. Turkish grammar has several derivational suffixes at its disposal. To mention only one, the suffix *-lI* is used to form an adjective on the basis of a noun, and its general meaning

is 'with'. For instance, *gözlük-lü adam* means 'man with glasses (on)'. The adjective *gözlük-lü* 'with glasses' is a derived predicate of which it can be assumed that it is the output of a Derived Predicate Formation rule. This rule accepts nouns as input: the input noun for *gözlük-lü* is the lexical form *gözlüQ* 'glasses'. When the latter form is taken as (the second) input predicate on a Compound Formation Rule together with *güneş* 'sun', the unit *güneş_gözlüQ-* 'sun glasses' will be the result, fully in line with the proposed model. Now, if not *güneş_gözlüQ-* but *güneş_gözlüğ-ü* (containing both the CM and the *ğ* as a result of a (morpho)phonological rule) were the output, we would again be in serious trouble when adjectival forms like *güneş_göz-lük-lü* are to be accounted for. Inputting *güneş_gözlüğ-ü* to a derivational rule for adjectives implies that the CM should be removed (and thrown away) and that the *ğ* should be restored to *k*. As has been indicated throughout, such operations are not compatible with an FG-view. Moreover, accepting that *güneş_gözlüQ-* is the basic compound form will lead to a more efficient handling of several phonological rules, since all rules are effected on the level of term expression. Of course, the difference between *koyun_baş-ı* 'sheep head-CM' and *koyun_baş-lı* 'mutton headed' can be explained in a quite similar way.

5. On Lexicalization

As has been indicated in the introduction of this paper, Lewis (1967) regards the constructions discussed so far as 'indefinite izafet' constructions. According to his judgment (or taste?), 'true' nominal compounds are those which consist of 'two nouns juxtaposed' (Lewis 1967:231 ff; 1983:39 ff) or which are based on an 'izafet group'. The main criterion, however, is that these compounds are written as one word (that is, without a space between the constituting elements), and that, in most cases, the elements composing it do not retain their primary meaning. Some examples are:

(27)	a.	*elbirliği*	(< *el* + *birlik* + CM)
		'co-operation'	hand union
	b.	*denizaltı*	(< *deniz* + *alt* + CM)
		'submarine'	sea underside

c *babaanne* (< *baba* + *anne* + CM)
'paternal grandmother' father mother

A characteristic feature of such compounds is that they represent different degrees of lexicalization, a fact that partially determines their syntactic properties. The (presumably) earliest stages of the compounds of (27) are given between the brackets. In the next stage of their developmental history, such constructions are no longer considered as 'free' constructs, but as 'ready made' predicates. This is the first step in the lexicalization process. As a result, the CM dissappears in three stages:

Stage 1: In (27a) the CM is still regarded as a compound marker. On the basis of case marking[2], it can be proven that the CM is not yet fully (but partially) incorporated into the predicate. If the CM were regarded a full part of the predicate, we might have expected the form *elbirliği-yi* 'co-operation-CM-ACC' (cf *denizaltı-yı* 'submarine-ACC'). However, we get: *elbirliğ-i-ni* 'co-operation-CM-ACC'. This implies that the compound is lexicalized as the predicate *el_birliQ-*, and that the CM is generated when the predicate is developed into a term.

Stage 2: In example (27b), the CM is no longer regarded as a marker, but as a part of the predicate. That is why the plural suffix, possessive or case markers are attached directly to the predicate (that is, directly to the former CM) e.g. plural: *denizaltı-lar* 'submarine-plur' (**denizalt-lar-ı*) 'submarine-plur-CM'; locative: *denizaltı-da* (**denizaltı-nda*) 'submarine-LOC', and also dative: *denizaltı-ya* (**denizaltı-na*) 'submarine-DAT'. Such compounds are stored in the lexicon as fossilized terms, containing the CM.

Stage 3: Predicates containing the original CM may eventually lose it again, as is the case in the example of (27c). Especially geographic names are viable to lexicalization processes and the effects thereof. A 1936 map of Istanbul, for instance, gives *Top Kapısı* ('gun' + 'gate' + CM) for the current *Topkapı*. In those days *Top Kapı-sı-na*, instead of modern *Topkapı-ya*, was used for 'to Topkapi'. Furthermore, we find nowadays names such as *Kadıköy* (formerly: *Kadıköyü*, < *kadı* 'judge' + *köy* 'village' + CM), *Bakkal Sokak* (< *Bakkal Sokağı* < *bakkal* + *sokaQ* 'street' + CM) and the like. Analogous to the development of *teldolap*[3] 'screen safe' (< *tel* 'wire, mesh' + *dolap* 'cupboard'), we may expect that a word like *buz dolab-ı* (which is also written as *buzdolabı*) 'refrigerator' will eventually evolve into *buzdolap*. This type of compound is stored in the lexicon as a frozen term, not containing the CM.

The fact that the CM vanishes after a period of time, means that the predicates involved are no longer regarded as compounds, but rather as bare predicates.

Although it seems as if *Bakkal Sokağı* becomes *Bakkal Sokak* again on the basis of 'reversed grammaticalization' (devoicing), this is true for the 'nominative' form only. Assuming, however, that its underlying form remains after the CM has disappeared (e.g. *Bakkal SokaQ*), all other (possessive and case) forms, as for instance *Bakkal Sokağ-a* 'to Grocer Street', are easily accounted for in terms of expression rules.

6. Summary and Conclusions

In designing a generative model for nominal compounds, two strongly interrelated problems had to be resolved:

(28) a. the overall structure of a Compound Formation Rule (CFR);
b. the status of the Compound Marker (CM).

To start with (28b), it was shown in section 1 that considering the CM as a part of the compounding process leads to various problems.

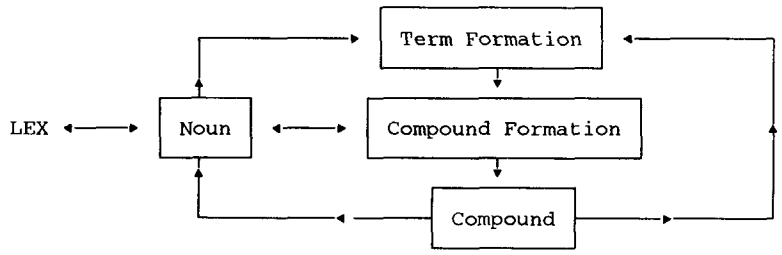
First, if the CM is generated as a part of a nominal compound, its form is determined by phonological properties of that compound. For instance, a compound ending in a vowel will trigger the CM *-sI* (as in: *misafir_odası* 'guest room'), and a compound ending in a consonant requires *-I* (as in: *çay_ev-i* 'tea house'). When such a compound is to be expressed in its plural form, the grammar should be able to either anticipate these circumstances (look ahead) or to carry out a repair on an (unnecessarily) ill-formed morpheme sequence afterwards. The latter operation implies that the original CM should first be deleted, that the plural marker should be attached, and finally, the CM should appear in an adapted form, e.g. *misafir_oda-lar-ı* and *çay_ev-ler-i*. Taking the point of view that no grammatical material may be deleted, transformed, or replaced, operations such as these should be avoided and an alternative description should be offered. Secondly, similar (deletion) operations would be necessary if a compound containing a CM as a product of its formation is used in a possessive construction. For instance, when *at_araba-sı* 'horse cart' is considered to be a ready-made compound on the basis of which *at_araba-mız* 'our horse cart' has to be formed, a deletion rule to suppress the CM must be assumed. An alternative approach is to regard the CM as a morpheme that is generated by a special rule of the (morphological) grammar. The CM can only occur when the nominal compound predicate is expressed as a term. In this approach, the output of

a compound formation rule is a bare nominal compound, which allows for a linear generation of forms based on it. Taking *at_araba* 'horse cart' as the output of a formation rule that does not attach a CM, this predicate can only be expanded by grammatical material via application of various expression rules. Hence, we get for instance: *at_araba-sı* '(his/her) horse cart'; *at_araba-mız* 'our horse cart'; *at_araba-lar-ı* '(his/her) horse carts'; versus: *at_araba-ları* 'their horse carts'; and *at_araba-lar-ımız* 'our horse carts' as the output of expression rules. In section 4, the attractiveness of this view was supported by an examination of nominal derivation. It was shown that a bare nominal compound predicate such as *güneş_gözlüQ-* can function as the input to a Formation Rule for Derived Predicates producing adjectives: *güneş_gözlük-lü*. All these examples clearly indicate that the CM is expressed on term level, instead of at some stage during the formation process of nominal compounds. Finally, in section 5 it was shown that this view is fully compatible with the three stages of compound lexicalization that can be distinguished.

As for (28a), the overall structure of a Compound Formation Rule, it was determined in section 2 that such a rule should be capable of generating compounds that contain a Cardinal Number, a Plural Marker, Adjectives or Proper Names. The best way to accommodate these requirements is to assume that the CFR takes a (fully specified) term as (a part of) its input. This assumption presupposes the processes of Term Formation and Term Expression. Input term structures need to be correctly layered during term formation so that expression rules will attach the CM on fully specified terms only. In the sections 2 and 3 we have seen that an input term may be based on a (basic) noun taken from the lexicon; it may be a ready-made term present in the lexicon (proper name); or it may be based on a compound predicate. This latter possibility is responsable for the Left Branching recursion (compound term plus nominal predicate). With respect to the second input of a CFR, this is always a bare nominal predicate; either it is a basic predicate taken from the lexicon, or it is itself a compound predicate. Taking a compound predicate as the secondary input of a CFR leads to Right Branching recursion (term plus compound nominal predicate). Of course, recycling of terms back into term formation must occur before expression.

The overall structure of the Compound Formation Process can be given as:

(29)



Since this model is based on the idea that fully specified terms function as the primary and nouns as the secondary input, it covers not only all kinds of 'normal' nominal compounds, but it is powerful enough to deal with coordinated terms as well, as is shown in a more elaborated treatment of Turkish compounds (Van Schaaik, 1989). Due to limited space, it was not possible to go into this topic in the present paper.

Notes

1. In example (2b), POSS denotes 'possessive suffix'.
2. An analysis based on possessive expressions containing the possessive suffix first person singular or plural is always somewhat fuzzy, since these suffixes begin with a vowel after consonant stems and with an *m* after vowel stems. For instance, *elbirliğimiz* 'our co-operation' could then be glossed as either *elbirliğ-imiz* 'co-operation POSS1s' or as *elbirliğ-i-miz* 'co-operation-CM-POSS1s'. In the latter case, the CM should be regarded as an integrated part of the compound.
3. According to Redhouse (1983:384) a *teldolap* is 'a small kitchen cupboard all the sides of which are made of screening'.

References

Baskakov, N.A. 1975. *Istoriko-tipologičeskaja xarakteristika struktury tjurkskix jazykov.* Moskva: Izdatel'stvo 'Nauka'.

Dik, S.C. 1978. *Functional Grammar.* Amsterdam: North-Holland.

Dik, S.C. 1980. *Studies in Functional Grammar.* London: Academic Press.

Dik, S.C. (ed.) 1985. *Valentie in Funktionele Grammatika, Tijdschrift voor Taal en Tekstwetenschap* 5:2.

Dik, S.C. 1989. *The Theory of Functional Grammar.* Part 1: *The Structure of the Clause.* Dordrecht: Foris.

Gadžieva, N.Z. 1973. *Osnovnye puti razvitija sintaksičeskoj struktury tjurkskix jazykov.* Moskva: Izdatel'stvo 'Nauka'.

Groot, C. de. 1987. "Predicate formation in Functional Grammar", *Working Papers in Functional Grammar* 20.

MacKenzie, L. 1987. "The representation of nominal predicates in the fund", *Working Papers in Functional Grammar* 25.

Lewis, G.L. 1967. *Turkish Grammar.* Oxford: Clarendon Press.

Lewis, G.L. 1983. *Teach Yourself Turkish.* Sevenoaks: Hodder & Stoughton.

Redhouse, 1983, *Çağdaş Türkçe-İngilizce Sözlüğü.* Istanbul: Redhouse Yayınevi.

Schaaik, G.J. van. 1985. "Valentie-reduktie in het Turks". *Tijdschrift voor Taal en Tekstwetenschap* 5:2, 127-39.

Schaaik, G.J. van. 1989. "The Generation of Turkish Nominal Compounds". Paper. Dept. of Computational Linguistics, University of Amsterdam.