Nanosyntactic Analysis of Turkish Case System

[Authors - to be added after review]

Aim. This paper takes two challenging characteristics of Turkish case system and shows that a nanosyntactic analysis is able to cover both of them. **Puzzle #1.** Some cases in Turkish show alternations between specific and non-specific forms, while other cases don't. Specifically, Turkish is well known for exhibiting differential object marking (NP-ACC vs NP_{bare}) and differential subject marking in embedded clauses (NP-GEN vs NP_{bare}). Both alternations are illustrated in the following example, which shows that both GEN and ACC serve the same purpose: conveying specificity. In the lack of these case markers, both the word 'thief' and 'home' would be non-specific.

(1) adam-**ın/Ø** gel-diğ-in-i ...
man-GEN/Ø come-NMLZ-POSS-ACC ...
'that a/the man came . . . '

(2) Dün doktor-**u/Ø** gör-dü-m.
yesterday doctor-ACC/Ø see-PST-1SG
'I saw a/the doctor yesterday.'

Interestingly, other cases do not show such an alternation, and we want to explain why. For example, both a specific and a non-specific dative look the same and have the ending *-a*. **Puzzle #2** concerns containment relations in morphology. In Caha (2009), it is proposed that cases stand in a containment relation. Specifically, the NOM case is the least marked case (characterized by just one feature K1). The nominative is contained in the accusative (K1+K2), which is in turn contained in the genitive (K1+K2+K3). The whole sequence proposed by Caha (2009) is in (3).

(3) $[_{COM}$ K6 $[_{INS}$ K5 $[_{DAT}$ K4 $[_{GEN}$ K3 $[_{ACC}$ K2 $[_{NOM}$ K1 [NP]]]]]]]

The hierarchy proposed by Caha is motivated by case syncretism (only adjacent forms may be syncretic) and by case containment (more marked cases contain less marked cases). Such syncretism/containment relations are illustrated in the first three columns of the table (examples from Caha 2011, representing Estonian, Tocharian and Vlax Romani). Of particular relevance is the fact that in all these paradigms, the ACC form serves as the foundation of the oblique cases. The oblique cases in some languages serve multiple functions as indicated by the brackets on the left.

CASE	church, SG, EST	horse, SG, TOCH	boy, SG, ROM	man, SG, TR
NOM	kirik	yakw-i	čhav-ó	adam
ACC	kirik-**u**	yakw-**em**	čhav-**és**	adam-**ı**
GEN	kirik-**u**	yakw-**em**-ts	čhav-**és**-k(or)o	adam-**ın**
DAT (ALL)	kirik-**u**-le	yakw-**em**-ts	čhav-**és**-ke	adam -a
INS (COM)	kirik-**u**-l	yakw-**em**-mpa	čhav-**és**-ar	adam -la

The puzzle is that in Turkish, the morphological containment holds only for ACC and GEN, but not for ACC and the other obliques. E.g., the comparison leads us to expect that the INS in Turkish could be **adam-ı-la*, with the ACC marker to the left of *la*. Interestingly, this expectation fails precisely in those cases which do not distinguish specific and non-specific forms.

Proposal, briefly. We propose a solution to both of these puzzles within Nanosyntax. The main idea is that Turkish nouns and cases can be composed into smaller, submorphemic features, and that the morphological realization of these features follows the phrasal spell-out mechanisms specified in Starke (2018). Our analysis is summarized in the table below. The table has two major compartments: the top part depicts the non-specific forms, the lower part depicts the specific forms. The surface forms are on the very left, and their analysis is to their right. The analysis has two parts. One part is identifying the features that individual forms contain (these are in the header of the table). Another part is determining how they are pronounced (this is indicated by shading).

	SURFACE FORM	CONCEPT	SPEC	NOM (K1)	ACC (K2)	GEN (K3)	DAT (K4)
NON-SPEC	adam (NOM)	[adam]		∅			
	adam (ACC)	[adam]			∅		
	adam (GEN)	[adam]				∅	
	adama (DAT)	[adam]				-a	
SPEC	adam (NOM)		[adam]				
	adamı (ACC)		[adam]		-ı		
	adamın (GEN)		[adam]			-ın	
	adama (DAT)	[adam]				-a	

The main idea behind the analysis is that below Caha's (2009) case features (K1 for NOM, K1+K2 for ACC, etc.), we find a privative feature SPEC that determines whether the interpretation is specific or non-specific. Non-specific nouns lack this feature (marked by black color in the table), while specific nouns have it. This is what distinguishes the upper part of the table and the lower part.

Analysis details. Our analysis of the specific/non-specific distinction is based on the idea that the case markers themselves do not spell out the feature SPEC, unlike Öztürk (2005). If we proposed, for instance, that *-ı* spells out this feature, it would be very difficult to explain why we do not find the specific/non-specific distinction in the dative and instrumental, recall the impossible hypothetical form **adam-ı-la*. Instead, we propose that SPEC is spelled out by the root, and the case endings reflect this indirectly via the spellout algorithm proposed in Nanosyntax. A crucial part of this algorithm is a matching procedure based on the so-called Superset Principle (SP), given in (4).

(4) SP (Starke 2009): A lexical item matches syntactic structure iff it contains that structure.

In analysis, that the nominal root is lexically specified for the features SPEC and NOM, see (5a). The root can therefore spell out all these features on its own, see the first three lines in the (lower) specific paradigm. It can, however, spell out also various subsets, see the remaining rows. In the specific paradigm, the ACC feature is realized by *-i*, see (5b) for its entry. In the GEN row, we find an additional -n, see (5c) for the entry.

(5) a. adam ⇔ [NOM (K1) [SPEC [CONCEPT]]]
b. -ı ⇔ [ACC (K2)]
c. -(n)ın ⇔ [ACC (K2) [GEN (K3)]]

(6) a. Ø⇔ [GEN (K3) [ACC (K2) [NOM (K1)]]]
b. -a ⇔ [DAT (K4) [GEN (K3) [ACC (K2) [NOM (K1)]]]]

Puzzle #2. The analysis of DAT in the specific declension involves Backtracking (Starke 2018). Note first that the only way for the DAT feature K4 to be spelled out, we must use *-a* with the entry in (6b), because no other ending contains K4. In order for this ending to match a constituent containing K1-K4, the root must backtrack from spelling out K1, see the last row of the table. The ACC ending *-ı* is thereby eliminated on the surface. An analogous proposal extends to the INS.

Puzzle #1. The distinction between the specific and non-specific declension emerges as a result of matching based on the Superset Principle (3). In the NOM of the non-specific paradigm, the root cannot spell out the K1 feature, because the syntactic constituent [[concept]nom] is not a subconstituent of the root's entry in (5a). Therefore, the root only spells out the concept part, and the zero non-specific ending spells out the NOM feature K1. The entry we assume for the Ø ending is in (6a). This entry allows it to also spell out the non-specific ACC and GEN; recall from (1) that these are Ø marked. In the DAT, the Ø ending no-longer matches all the case features (it does not contain them), and therefore, the ending *-a* is used. This way, a difference between specific and non-specific forms is found only in ACC/GEN, but not the other cases.