Leipzig Spring School on Linguistic Diversity: MPI-EVA
Jaklin Kornfilt, Syracuse University and MPI-EVA: **Topics in Turkic Syntax**
kornfilt@syr.edu **26-28 March, 2008**

RCs: "the precise relation between the relative clause and its head is not specified by the syntactic construction itself". Johanson (1998) "The Structure of Turkic"

To what extent true:
When target can be one of a number of arguments, e.g. either dative or ablative (typically):

Turkish:

(1) taşın-dığ-ım ev
move-FN-1.SG house
'the house I moved to/from'

This is because the verb taşın 'move' can take either a dative or an ablative complement:

(2) bu ev -e / ev – den taşın – dı – m
this house -DAT / house-ABL move-PST-1.SG
'I moved to/from this house'

Given that typically, in all Altaic languages (perhaps more generally, in all verb-final languages) simple RCs have gaps rather than relative pronouns as RC targets, this ambiguity is typical and systematic for just those verbs which allow different options for complements with a range of cases, as in this example. A similar example from Kirghiz, also from Johanson (1998):

(3) men kel -gen ša:r
I come -participle city
'the city I have come to/from'

But this is a limited type of ambiguity and not of vagueness. It's limited to certain verbs, and to certain possible arguments of such verbs.

It's true that some types of RCs don't necessarily express the agent; one type involves passive modifying clauses:

(4) dondurma sat -ıl -an yer
ice-cream sell -PASS -participle place
'The/a place where ice-cream is sold'

Because the agent is missing, the entire construction is impersonal. But the same is true of any passive construction, not just of the passive in a relative clause. An agent phrase can be provided in a passive RC, just as it can in any passive construction:

(5) başbakan tarafından dondurma sat -ıl -an yer
prime minister by ice-cream sell -PASS -participle place
'The/a place where ice-cream is sold by the Prime Minister'

Yet another type of impersonal RC is a type that corresponds to infinitival RCs in IE-languages; in Turkish, the future tense marker is used in that function, and no subject/agent can be expressed overtly:

(6) [PRO oku-yacak] bir kitap
read-FUT a book
'a book to read'

Other Turkic languages have corresponding RCs; cf. Johanson (1998):

Kirghiz:

(7) oqu -r kitep
read -AOR book
'A book to read'

Bashkir:

(8) uqı-hı kitap
read-AOR book
'A book to read'

The impersonal nature of this construction stems from the infinitive-like character of the predicate and not from the RC per se. Infinitives don't allow for any overt subject in any of these languages; cross-linguistically, overt subjects are unusual in general, as typological studies have shown. Notice that in Turkish, in RCs that target a non-subject, the predicate has a subject agreement marker, even if the subject is "silent" but is understood as a referential pronoun, i.e. as *pro*:

(9) oku-yacağ -ım bir kitap
read-FUT -1.SG a book
'a book which I will read'

But in (6), the predicate has no agreement marker, just as with infinitives in general:

(10) [PRO kitab -ı oku -mak] isti -yor -um
book-ACC read -inf want-PRES.PROG.-1.SG
'I want to read the book'

The main difference between the Turkic "infinitival" relatives and, say, English infinitival relatives is that morphologically, the nominalized predicate is not an infinitive, but is a participle of some sort, depending on the individual language. This is because infinitives (as well as some other predicate forms, such as subjunctives) don't allow for any "operator — variable" constructions, i.e. relative clauses and embedded interrogatives; let's claim that clauses with such predicates aren't "large" or "high" enough to house such operators (corresponding to wh-pronouns in IE-languages). This is why these languages must use some other type of nominalized morphology, such as a factive nominalization, a factive participle, some aspectual participle and the like. From the point of view of the syntax, however, constructions such (6), (7), and (8) correspond to the English translation, i.e.:

(11) a book [PRO to read]

whereby PRO is interpreted as any human entity, with the possibility to pick up reference in the syntactic or discourse context in which the construction appears.

We thus see that in general, the relationship between a relative clause and its head is rather narrowly specified in Turkic languages (remember the examples in previous handouts and discussions), almost as narrowly as is the corresponding relationship is in IE-languages. The instances where the relationship is less specified are systematic and limited to certain ambiguities, and there is no real vagueness.

Vagueness may arise in other constructions, such as in certain N-complement constructions:

Karachay-Balkar:

(12) [et biš -gän] iyis
meat cook -RelPart smell
'The smell of meat cooking'

Sakha:

(13) [[üle bul-ar -ga yaraxan] fizika]
work find-AOR-DAT difficult physics
'Physics, which is hard to get a job in'

The Sakha example can be interpreted as an RC. But even with examples such as these, it is possible to make some formal sense. In (13), the following analysis has been suggested by a Sakha linguist, Dr. Nadya Vinokurova, who proposes that the next example is the source of the previous RC:

(14)[Fizika [üle bul-ar -ga] yaraxan]
physics work find-AOR-DAT difficult
'Physics is difficult to find a job in'

NV suggests that the embedded clause is an "infinitival" clause (with a PRO subject of the sort we posited for Turkish infinitivals). (13) would be the result of relativization, targeting the subject of (14). A similar pair of related constructions, but without a further embedding, would be as follows:

(15) a. Bu sorudax Masha-qa cepceki
this assignment Masha- DAT easy
'This assignment is easy for Masha'

b. [e_i Masha-qa cepceki] sorudax$_i$
Masha-DAT easy assignment
'A/the assignment which is easy for Masha'

As for the Karachay-Balkar example, we know that N-complement clauses in Turkish and other Turkic languages consist of compound-like constructions, at least for certain nominalization types. Thus, what's involved there is probably the formation of a nominal compound such as 'cooking-smell', which is them predicated of 'meat' and could thus be predicated of other appropriate entities. While the relationship between *meat* on the one hand and *smell* may be vague potentially, the nominalized predicate narrows it down. Note also that even in IE-languages, nominalizations introduce a measure of ambiguity; e.g. *the teaching of the students* could be interpreted with *students* as the theme as well as the agent, and this ambiguity can be disambiguated in context.

It appears, then, that the relationship between a modifying clause in RCs, and its head is not vague in Turkic languages, but rather ambiguous in predictable ways; the relationship between a complement clause and its nominal head may likewise be no more vague than its corresponding constructions in IE-languages.

What about embedding in general? We have seen head-final structures: N-complement clauses and RCs with the nominal head after the clause, and some embedded clauses, with the verb after the clause. This is to be expected in generally head-final languages. Typically, such embeddings are nominalized in some fashion. (There have been typological and theoretical attempts to find explanations for this, but none of them are satisfactory. Also, Turkic languages have clear examples showing that while this is a general tendency, it's not a universal; there are head-final clausal embeddings that are fully finite, i.e. show the entire range of root tense and aspect markings, with nominative rather than genitive subjects.) For example, Turkish:

(16) [Sen geç gel -ecek -sin] san -ıyor -du -m
you [Nominative] late come -FUT -2.SG believe-PRES.PROG. -PST-1.SG
'I was thinking (that) you would come late'

It has often been claimed that such constructions are borrowed constructions—borrowed as structures, i.e. finite constructions, from other languages, usually IE-languages, albeit with native lexical and morphological items. This may be plausible for head-initial structures, but not for these head-final ones. It is more plausible to view these as native, and the result either of native historical

development, or as a structure existent even in older stages. The latter doesn't seem to be the case, at least as far as the (limited) documentation shows; but there is no evidence against the former.

What about head-initial finite constructions? There, it makes more sense to posit borrowing of the syntactic structure, especially in versions with overt complementizers such as *ki*, an item borrowed from Persian:

(17) san -ıyor -du -m ki [sen geç gel -ecek -sin]
believe-PRES.PROG.-PST-1.SG that you [Nominative] late come -FUT -2.SG
'I was thinking (that) you would come late'

Note that leaving out the complementizer is possible, leading to fully finite embedded clauses that follow the matrix verb directly; however, such structures are more often found preceding the verb rather than following it; when they follow the matrix verb, they more often show up as in (17), i.e. with the Persian complementizer.

Johanson (1998) has an interesting view of examples such as (17): "Though such clauses may resemble English constituent and relative clauses, they often differ considerably from them. The junctor signals a close semantic connection with the subsequent clause, but normally does not introduce it, e.g. Turkish *Sanırım ki, gelecek* 'I think (s)he will come', Uzbek *Men bildim ki, siz kelgan ediŋiz* 'I got to know that you had come'. Clauses of this type are mostly not embedded as postpredicative constituents of the preceding clause, but appended to it in a looser way."

There are two claims here: 1. That the "junctor", i.e. the complementizer, does not form a constituent with the embedded clause, and 2. That the embedded clause is not a constituent of the preceding, i.e. matrix, clause, but is "appended to it in a looser way".

The first claim isn't very strong; the complementizer may still form a *syntactic* constituent, but be *phonologically* cliticized to the preceding phonological word, which is typically the predicate of the matrix clause. Given that in all of these languages, clitic elements are postclitics, this is a plausible view. Thus, the commas set by Johanson in his examples above would express *phonological* sequencing, but not *syntactic*, phrasal constituent boundaries.

As for the second claim, it is hard to evaluate it without a clearer view of the loose connection referred to, or any evidence for such a connection, contrasting with a more familiar, tighter kind of connection. Johanson is probably right that this is so in RCs of the head-initial kind; e.g. in Turkish:

(18) Bir baba ki çocuğ -un -u döv -er kötü bir baba -dır
a father that child -3.SG -ACC hit -AOR bad a father -is
'A father who hits his child is a bad father'

Such RCs, at least in Turkish, are actually parentheticals rather than genuine embedded clauses that modify (in the sense of strict modification) their head. They are restricted in terms of tense/aspect of the (finite) embedded predicate, for example. As a matter of fact, colloquial, modern Turkish uses these structures less and less. Thus, Johanson's statement does fit such right-branching, finite RCs.

But his statement has lesser validity for general finite embeddings with *ki*, which are much less restricted (if at all) and which are still rather widely used.

There is a Turkic lexical item that resembles *ki*: this is the lexical item *kim*, which is documented in Old Turkic and is an interrogative element, used also as an indefinite pronoun of the kind found in IE languages, e.g. *whoever*. An example from OT (Erdal 1998):

(19) **kim** qayu küseser qutlu • tınlı • lar ara etizü olorup aša • alı, **ol kiši** edgü qılınč qılzun
whoever that person
'Whoever wishes to enjoy sitting among the blessed creatures making music, that person should carry out good deeds'

This is a correlative construction, where the embedded clause with the wh-element is a conditional, and where it does not form a constituent with any head. Instead, the reference of the wh-element is picked up by an independent NP, here: *ol kiši*, in the matrix. This is a pattern seen in some IE-languages, too, especially among languages of the Indian subcontinent, e.g. Hindi.

Of course, this kind of pattern with a "finite" embedding may have been borrowed even during OT-times. Nonetheless, the existence of such examples in OT shows that fully finite embedded clauses is not a late borrowing from non-Turkic languages.

Finite embeddings without a complementizer that precede the matrix verb are found in early Ottoman, too, e.g. in the 14th century (cf. Kerslake 1998):

(20) Ya: Ömer,[bu özr -i Taŋrı sen-den qabu:l eyle-sin] bil -me -z -di -m
Oh Umar this excuse-ACC God you-ABL acceptance do-OPT.3 know-NEG-NEG.AOR-PST-1.SG
'Umar, I did not think that God would accept this excuse from you'

It is possible that this pattern was borrowed from Persian, given that Persian culture was prestigious and was emulated by the Turks of Anatolia during the period in question. However, one would need to establish that such pre-verbal finite clauses were indeed used widely in the Persian of that period.

There is no doubt that in some modern Turkic languages, some IE-features have indeed been borrowed; e.g. Gagauz must have borrowed wh-movement from Slavic; cf. Menz (1999):

(21) Nere-dæn "gagauz" ad -ı gel-di
where-ABL Gagauz name-CMPD.MRK come-PST
'From where did the name "Gagauz" come?'

While Turkish, as a closely related language, is very much word-order free, to the extent that it can move around wh-elements otherwise preferred in immediate pre-verbal position, the order in (21) with the wh-element in sentence-initial topic position, preceding the subject, is very marked. For such an order to be well-formed, the discourse or the pragmatic setting must be rather special, given that wh-elements are focused by default and thus hard to topicalize in general. In Gagauz, on the other hand, sentence-initial order of wh-elements is found predominantly. Thus, it appears that Gagauz has

borrowed wh-movement of appropriate wh-constituents to the specifier position of CP, a typical IE-property.

References:

Erdal, M. (1998) "Old Turkic"; in Johanson & Csató; 138-157.
Johanson, L. & É. Á. Csató (1998) *The Turkic Languages*; London & New York: Routledge
Johanson, L. (1998) "The Structure of Turkic"; in Johanson & Csató; 30 – 66.
Kerslake, C. (1998) "Ottoman Turkish"; in Johanson & Csató; 179-202.
Menz, A. (1999) *Gagausische Syntax: Eine Studie zum kontaktinduzierten Sprachwandel*; Wiesbaden: Harrassowitz