The processing of relative clause attachment ambiguities in Turkish

Bilal Kırkıcı

Kırkıcı, Bilal 2004. The processing of relative clause attachment ambiguities in Turkish. *Turkic Languages* 8, ###-###.

The aim of this study is to investigate the way native speakers of Turkish resolve relative clause (RC) attachment ambiguities in sentences which contain a relative clause followed by a complex noun phrase with two potential attachment sites. The structures under investigation are relative clauses followed by complex NPs with genitive constructions [$NP1_{GEN}$+NP2] as in (1) and relative clauses followed by complex NPs containing postpositional phrases [[NP1 $P]_{PP}$+NP2] as in (2): (1) *ABD'de yaşayan aktörün oğlu dün uçak kazası geçirdi*, (2) *Elinde çiçek olan görevlinin yanındaki adam benim fizik öğretmenim*.

Previously conducted studies have found cross-linguistic differences in the resolution of such structural ambiguities, showing that in some languages the RC is predominantly attached to the first NP, i.e. *aktör*, while speakers of other languages attach the RC to the second NP, i.e. *oğlu*. However, it has also been found that lexical-semantic information like the presence of a pre-/postposition in the complex NP as in (2) and the animacy features of the potential hosts constitute determining factors in the choice of the NP which the RC will be attached to.

The results of two off-line (untimed) questionnaire experiments conducted with adult native speakers of Turkish reveal that sentence comprehension and (RC) ambiguity resolution is highly influenced by the semantic features of the constituents involved and speak against a parsing account for Turkish based on syntactic or locality-based constraints alone.[*]

Bilal Kırkıcı, Middle East Technical University, Faculty of Education, Dept. of Foreign Language Education, 06531 Ankara, Turkey.

[*] This is a revised version of a paper delivered at the 12th International Conference on Turkish Linguistics (ICTL 2004, Dokuz Eylül University, Izmir).

1. Introduction

A longstanding question in psycholinguistic research is whether the human language processor employs a universal parsing strategy in sentence processing or whether different languages (or groups of languages) make use of different processing strategies. A substantial amount of cross-linguistic research into the question of the 'universality of parsing strategies' has focused on the processing of relative clause (RC) attachment ambiguities as in the famous example (1) below, in which the RC can be attached either high, to NP1 (*the servant*), or low, to NP2 (*the actress*).

(1) *Someone shot* [*the servant*]$_{NP1}$ *of* [*the actress*]$_{NP2}$ *who was on the balcony.*
HIGH LOW

For many years, the dominant view was based upon the Recency (Gibson et al. 1996), Late Closure (Frazier & Fodor 1978) or Right Attachment (Kimball 1973) principles, which basically contend that regardless of cross-linguistic lexical and grammatical differences, all human languages are processed by means of the same mental machinery according to which constituents such as RC modifiers are attached to the last (or closest) potential constituent (Baccino & de Vincenzi & Job 2000, Fernandez 2003). Underlying the dominant view was a principle of economy, where the parser takes the 'least effort' option of attaching a constituent to the most recently processed (or closest) phrase, regardless of the language being processed.

2. Cross-linguistic differences

The earliest study that challenged the 'universalist' view was that by Cuetos & Mitchell (1988), who found that while English subjects in questionnaire studies displayed a Late Closure trend and tended to attach RCs in constructions comparable to (1) to the lower NP, i.e. to *actress*, speakers of Spanish did not follow Late Closure but showed a high attachment preference instead. This led Cuetos and Mitchell to the conclusion that certain parsing strategies may not be universal but subject to cross-linguistic variation and may have to be learnt by experience (Roberts, 2003).

Following Cuetos & Mitchell, further studies examining RC attachment preferences with two-site NPs incorporating genitives in other languages have provided additional support for the view that the Late Closure principle may not be generalised cross-linguistically. While a low-attachment preference as predicted by the Late Closure theory has been found for some further languages like Arabic, Brazilian Portuguese, Romanian, Norwegian and Swedish, a high-attachment preference has been attested for Greek, Dutch, Afrikaans, Russian, Croatian, Polish, Spanish, French and German (Fernandez 2003). These findings have pointed to the possibility that certain parsing strategies are language-specific rather than universal and that strategies other than Late Closure may be operative in different languages.

A number of theoretical and experimental attempts have been made to explain these cross-linguistic differences. However, since a thorough discussion of all established accounts would be beyond the scope of this study,[1] only the Tuning hypothesis (e.g., Mitchell et al. 1995), the Predicate Proximity theory (Gibson et al. 1996) and the Construal theory (Frazier & Clifton 1996), which are particularly relevant to the properties of Turkish and to the results obtained in the present study, will be shortly discussed below. According to the Tuning hypothesis, the human language processor makes attachment decisions in temporarily ambiguous conditions on the basis of the frequencies of attachment preferences in unambiguous conditions. That is, in a language in which unambiguous RCs are typically attached high, the tendency to attach an ambiguous RC high will be stronger than a tendency to follow low-attachment, and vice versa. Mitchell et al. (1995) have reported findings in support of the Tuning hypothesis that display positive correlations between the frequency distributions of attachments obtained from corpus data and RC attachment preferences obtained from experimental studies. See, however, other studies (e.g., Gibson & Schütze 1999, Mitchell & Brysbaert 1998) for conflicting evidence.

Gibson et al. (1996) try to account for the cross-linguistic variations found by proposing a principle they call Predicate Proximity, which supposedly interacts with the universal Recency principle. According to the Predicate Proximity principle, ambiguous modifiers will be attached to constituents as structurally close as possible to the main predicate of the sentence. It is hypothesised that in non-configurational languages (i.e., languages with a relatively freer word order) like German, Greek and Turkish, Predicate Proximity is able to outrank the universal Recency strategy and results in a high-attachment preference (i.e., attachment to *servant* in (1) above), whereas in configurational languages like English the Recency strategy is more dominant and results in low-attachment instead. Support for Predicate Proximity has been found for a number of non-configurational languages like Greek, German and Spanish (Papadopoulou & Clahsen 2002).

According to Frazier & Clifton's (1996) Construal theory, the resolution of ambiguities in modifier attachment is primarily implemented on the basis of lexical-semantic rather than syntax or locality-based information alone. The point of departure of Construal is a hypothesised distinction between primary and non-primary phrases[2] and the assertion that non-primary phrases such as RCs are associated with the closest thematic processing domainthe extended maximal projection of the last theta-assigner. In a sentence with a complex genitive host like

[1] See Fernandez (2003) for an excellent and thorough discussion of almost all relevant mainstream theories of RC attachment processing.

[2] Frazier & Clifton (1996) define primary phrases as comprising subject and main predicates, and complements and obligatory constituents of primary phrases, while non-primary phrases are non-obligatory constituents like RCs, adjunct predicates and phrases related by conjunction.

(1) above, the closest thematic processing domain would be the entire complex NP *the servant of the actress* since the preposition *of* is not a theta-assigning preposition. Therefore, Construal would not predict any distinctive attachment preference in sentence (1), but would propose that "all possible hosts within this domain are evaluated in parallel using a range of information" (Frazier & Clifton 1996: 365) such as semantic/pragmatic information, prosody, the frequency of past exposure to a certain attachment pattern, and preferences imposed by locality-based parsing strategies such as Recency and Predicate Proximity (Felser et al. 2002).

(2) *Someone shot* [*the actress*]$_{NP1}$ *with* [*the servant*]$_{NP2}$ *who was on the balcony.*

In sentence (2) above, however, the complex NP *the actress with the servant* contains a thematic preposition *with*, which thus creates a local thematic domain of its own. Since the local thematic domain created by the preposition *with* is at the same time the closest thematic processing domain to the following RC, Construal would predict that in sentences like (2) the ambiguous RC should be associated with this thematic domain and therefore be attached low, i.e. to NP2 (*the servant)*, as in (3) below.

(3) *Someone shot the actress* [*with the servant who was on the balcony*].

This proposed preference to attach RCs low when NP2 is introduced by a thematic preposition has been confirmed cross-linguistically by a number of studies (e.g., Gilboy et al. 1995), even for languages like French, Greek and Spanish which display a high, i.e. NP1, attachment in sentences containing complex genitive NP hosts like sentence (1) above (Papadopoulou & Clahsen 2002).

3. The present study

The present study explores the RC attachment preferences of adult native speakers of Turkish, a language which has not been investigated as yet in the framework of RC attachment, by means of two off-line (untimed) questionnaires. It is important to note that in contrast to almost all languages that have been analysed hitherto, Turkish predominantly makes use of prerelative, rather than postrelative, constructions[3] as illustrated in (4), a Turkish equivalent of (1):

(4)	*Birileri,*	*balkon-da*	*dur-an*
	Someone-NOM	balcony-LOC	stand-REL

[3] See Kamide & Mitchell (1997) for an analysis of RC attachment preferences in Japanese, which also employs prerelative constructions.

[*aktris-in*]	[*hizmetçi-si-ni*]	*vur-du*.[4]
LOW	HIGH	
actress-GEN	servant-POS-ACC	shoot-PAST

Turkish prerelatives are non-finite, have a nominalised predicate with a special morphology, do not make use of complementisers (Aygen 2003), and employ a constituent order completely different from almost all languages that have been analysed thus far within this context, except for Japanese which displays a similar structure in RCs (e.g., Kamide & Mitchell 1997). As can be seen in (4), the RC (*balkonda duran)* precedes both NP-LOW (*aktris*) and NP-HIGH (*hizmetçi*) and, in addition, NP-LOW precedes NP-HIGH.

4. Method

4.1. Participants

The participants employed for the present study were students from various departments at Middle East Technical University, Ankara. All participants were native speakers of Turkish, reported to have normal or corrected-to-normal vision, were unaware to the purpose of the study and participated in the experiments on a voluntary basis. 48 participants (10 males, 38 females) took part in Experiment 1 and 42 (19 males, 23 females) in Experiment 2. None of the subjects participated in both experiments.

4.2. Materials

Experiment 1: [+human] Host NPs. For Experiment 1, a questionnaire in Turkish was designed that included 40 sentences in total. 14 of these sentences were ambiguous experimental items and 26 were filler items of various grammatical types to divert the participants' attention from the specific grammatical structure under investigation. Each experimental sentence was constructed in two different versions, containing a complex NP host with either a genitive construction as in (5a) or the postposition *yanında* 'next to' as in (5b). Both NPs were exclusively [+human], appeared in the singular and were matched for frequency to prevent possible frequency effects.[5]

[4] Turkish makes also use of postrelative constructions, though to a lesser extent. The postrelative equivalent of (4) would be *Birileri aktrisin hizmetçisini vuru, ki o balkonda duruyordu*, which employs the complementizer *ki*, a borrowing from Persian. See Underhill (1974) for further information.

[5] The experimental nouns were matched for frequency on the basis of the frequency count provided by Göz (2003). A frequency ratio of minimum 0.65 between each pair of nouns was used as a criterion.

(5a) *Şoför, şehir merkezinde oturan profesörün sekreterini gördü.*
(5b) *Şoför, şehir merkezinde oturan profesörün yanındaki sekreteri gördü.*

Two different versions of the questionnaire were constructed with each version containing one version of each experimental sentence only. Each questionnaire contained an equal number of sentences of type (5a) and (5b), which constituted the Genitive (GEN) and Postposition (PP) conditions, respectively, and appeared in randomised order.

Experiment 2: [-human] Host NPs. The questionnaire used in Experiment 2 comprised a total of 36 sentences (12 experimental items, 24 filler items). The experimental sentences used were similar to sentences of type (5a) in Experiment 1 in that they included singular nouns in the genitive condition, with the difference that both nouns in the complex host NP were exclusively concrete [-human] nouns as illustrated in (5c).[6]

(5c) *Yazar, parklarıyla ünlenen ülkenin bąkentini ayrıntısıyla anlattı.*

4.3 Procedure

In both experiments, participants were instructed to read each sentence and indicate as spontaneously as possible for each of the items which of the possible interpretations they considered most appropriate. See (6) for illustration.[7]

(6) *Yazar, parklarıyla ünlenen ülkenin bąkentini ayrıntısıyla anlattı.*
(a) *başkent parklarıyla ünlenmişti*
(b) *ülke parklarıyla ünlenmişti*

The order of appearance of the NPs in the choices was randomised so that participants would not develop answering strategies. The completion of each questionnaire took about 20 minutes.

5. Results

Table 1 presents the mean percentages of responses provided for the experimental sentences in Experiments 1 & 2.

Table 1: Mean percentages and standard deviations of responses.

	NP-LOW	SD	NP-HIGH	SD
Experiment 1				
GEN [+human]	55%	29.7	45%	28.9
PP [+human]	68%	26.2	32%	25.6

[6] Experiment 2 did not analyse complex host NPs containing a postposition as in (5b).
[7] Items in Experiment 1 contained three choices.

Experiment 2				
GEN [-human]	63%	24.2	37%	24

Experiment 1. As can be seen in Table 1, subjects in Experiment 1 showed a higher preference to attach the RC low in the PP condition (68%), but displayed no strong preference in the GEN condition (55% Low, 45% High). The low-attachment preference found in the PP condition was found to be statistically significant when analysed by subjects ($t1(47) = 4.983$, $p < 0.0001$) as well as by items ($t2(13) = 3.943$, $p < 0.005$). No significant difference between the high and low attachment responses, however, was found in the GEN condition when analysed by subjects ($t1(47) = 1.243$, $p = 0.220$) or items ($t2(13) = 1.144$, $p = 0.272$), indicating that the subjects had no real attachment preference in the GEN condition with [+human] nouns.

One-sample t-tests further showed that subjects responded at chance-level in the GEN condition ($p > 0.2$), whereas they responded significantly above chance in the PP condition ($t1(47) = 4.769$, $p < 0.0001$; $t2(13) = 3.827$, $p < 0.005$).

Experiment 2. As Table 1 displays, the subjects in Experiment 2 showed a preference for low attachment (63%) when the nouns in the GEN condition were [-human]. This obtained low-attachment preference was also supported by the statistical analyses, which revealed significant differences between high and low attachment responses by subjects ($t1(41) = 3.579$, $p < 0.005$) as well as by items ($t2(11) = 3.272$, $p < 0.01$). One-sample t-tests showed that this low-attachment preference was above chance-level ($t1(41) = 3.504$, $p < 0.005$; $t2(11) = 3.229$, $p < 0.01$).

In other words, in contrast to the [+human] GEN condition in Experiment 1, where no attachment preference was obtained, subjects did display a statistically significant low attachment preference when the nouns in the complex host NP were [-human] in the GEN condition.

6. Discussion

The results of the present study have shown that the native speakers of Turkish who have taken part in the present study:

(a) do not display any attachment preferences in sentences where an ambiguous RC has two potential [+human] attachment NP hosts in the genitive condition (Experiment 1)
(b) prefer to attach the ambiguous RC to the low NP in conditions where two potential [+human] NP hosts are joined by means of a postposition (Experiment 1)
(c) tend to choose the low NP as an attachment host when two [-human] NPs exist as potential attachment hosts in the genitive condition (Experiment 2).

Finding (b) is very much in line with the predictions of the Construal theory, which predicts that the presence of a theta-assigning pre-/postposition will create a comparatively stronger preference to associate the ambiguous RC with NP-LOW due

to the fact that the thematic processing domain created by the postposition *yanında* and including only NP-LOW is closest to the RC as illustrated in (7). As reported above, the subjects in the present study indeed showed a statistically significant low-attachment preference (68%) in the PP condition. In this sense, finding b) provides further support for the cross-linguistic finding that modifier ambiguity resolution is influenced by lexical-semantic information.

(7) *Şoför,* [*şehir merkezinde oturan* [[*profesörün*]$_{\text{NP-LOW}}$ *yanındaki*]]$_{\text{PP}}$ *sekreteri gördü.*

Findings (a) and (c) are related and need to be evaluated together. From the perspective of the Construal strategy, in the GEN condition (NP_{GEN}+NP) both NPs should be available as potential hosts since the thematic processing domain closest to the ambiguous RC is the entire complex NP. Therefore, the interaction of universal, locality-based parsing principles such as Recency or Predicate Proximity should determine the final attachment preference. Considering the fact that Turkish clusters with languages as Greek and German in that all are non-configurational languages allowing for free word order, it should be expected that Predicate Proximity outranks the Recency principle, resulting in the tendency to attach the relative clause to the NP closest to the main predicate (i.e., NP-HIGH) as found in other non-configurational languages.

Interestingly, however, the subjects in the present study displayed no attachment preference in the [+human] GEN condition (55% low-attachment, 45% high-attachment) and preferred to attach the ambiguous relative clause low in the [-human] GEN condition (63% low-attachment). In other words, although the syntactic structures were identical in (a) and (c), the change in lexical-semantic information [±human] conveyed through the NPs brought about a shift in the attachment preference observed. Thus, while the participants behaved in accordance with the Recency principle in the [-human] GEN condition, they did not apply a structural processing strategy in the [+human] condition. Unfortunately, none of the currently prevailing structural accounts of sentence processing can provide satisfactory explanations for this dissociation.

A potential explanation could be sought within the framework of the Tuning hypothesis. A similar dissociation between the attachment preferences for [+human/+human] vs. [-human/-human] GEN NP hosts as found in the present study was reported for Dutch speakers in two completion experiments (Desmet & Brysbaert & de Baecke 2002), which was found to correspond to the corpus frequencies obtained for Dutch[8] as predicted by the Tuning Hypothesis. Thus, a possible step in explaining the dissociation found in the Turkish data could be the

[8] Corpus frequencies: [+human/+human] low attachment: 33%, [-human/-human] low attachment: 70%; Experimental findings: [+human/+human] low attachment: 36%, [-human/-human] low attachment: 79% (Based on Desmet et al. 2002: 890, Table 4).

analysis of corpora to see whether such a dissociation can indeed be manifested in written/spoken sentence production in Turkish. It could indeed be the case that speakers of Turkish prefer to associate RCs with the low NP when producing (NP_{GEN}+NP) structures of the type [-human/-human] more frequently than with [+human/+human] NP heads. Needless to say, this is nothing more than a speculative approach to the present data that needs serious back-up from well-established corpus-data.

Another possible explanation for the results obtained in the [-human] GEN condition (c) could be established in line with the Gricean maxim of quantity, which requires that the speaker should be as informative as necessary, providing neither too much nor too little information.[9] In Turkish, it is possible to place the RC between the first and second NP in complex GEN heads as in (8) below, which forces the preceding RC to be taken as modifying NP2 (*başkent*). In this sense, the low-attachment preference found for [-human] complex genitives would be understandable in the Gricean framework since an unambiguous high-attachment enforcing structure exists in the language, which could have been employed but wasn't, and the subjects may therefore have taken the ambiguous structure as an indication of low-attachment instead.

(8) *Yazar,* [*ülkenin*]$_{NP1}$ *parklarıyla ünlenen* [*başkentini*]$_{NP2}$ *ayrıntısıyla anlattı.*

Then, however, the very same principle would also be expected to hold for [+human] GEN constructions, which can also be disambiguated by means of the same structure as shown in (9), which was not the case, though. Whether or not this dissociation is a reflection of a general tendency in Turkish to use the unambiguous structure predominantly with [-human] NP hosts is a possibility that may be evaluated in future research, but can not help shed further light on the present findings.

(9) *Şoför,* [*profesörün*]$_{NP1}$ *şehir merkezinde oturan* [*sekreterini*]$_{NP2}$ *gördü.*

7. Conclusion

The aim of this study was to constitute a first step in establishing the preferences of Turkish speakers in the processing of relative clause (RC) attachment ambiguities. As stated above, despite the fact that many and diverse languages have been analysed to date within this framework, contributing to the wider question of the universality of processing strategies, an important and widespread language like Turkish has unfortunately been rather neglected thus far.

[9] See Gilboy et al. (1995) for a similar account of results obtained in the GEN condition in English.

The results of two questionnaire experiments have shown that, overall, the Turkish parser appears to be highly sensitive to lexical-semantic information because changes in the lexical-semantic information conveyed through NPs ¯[human] had an important impact on the attachment preferences observed in the GEN condition and, similarly, the use of a lexical/thematic postposition like *yanında* led subjects to show a relatively strong low-attachment preference. The predictions of purely structure-based theories for Turkish, on the other hand, were not fully borne out by the findings obtained. It therefore seems that future research on RC attachment ambiguity in Turkish needs to set up expectations by also taking into consideration various aspects such as pragmatic constraints, lexical/semantic effects and frequency of exposure rather than entirely structural/locality-based factors.

Needless to say, the results of the present study can as such not be generalised given the fact that the subject population was homogenous in many respects (e.g., 100% university students; small age-range) and small in size and due to the rather mixed findings that were obtained. Nevertheless, in terms of the larger, cross-linguistic picture of RC attachment preference research, it is possible to say that the present study patterns with previous studies that have underscored the point that a universalist account based on entirely structural processing strategies, and eschewing alternative intervening factors, may not be maintained in light of the cross-linguistic findings attested.

Abbreviations

NOM nominative
LOC locative
REL 'subject relativizer' morpheme (Hankamer & Knecht, 1976)
GEN Genitive
POS possessive
ACC accusative
PAST past tense morpheme

References

Aygen, G. 2003. Are there "non-restrictive" prerelatives in Turkish? *Harvard Working Papers in Linguistics* 8, 199-212.

Baccino, T. & de Vincenzi, M. & Job, R. 2000. Cross-linguistic studies of the late closure strategy: French and Italian. In: De Nincenzi, M. & and Lombardo, V. (eds.) *Cross-linguistic perspectives on language processing*. Dordrecht: Kluwer Academic Publishers. 89-118.

Cuetos, F. & Mitchell, D. 1988. Cross-linguistic differences in parsing: restrictions on the use of the Late Closure strategy in Spanish. *Cognition* 30, 73-105.

Desmet, T. & Brysbaert, M. & de Baecke, C. 2002. The correspondence between sentence production and corpus frequencies in modifier attachment. *The Quarterly Journal of Experimental Psychology* 55A (3), 879-896.

Felser, C. & Roberts, L. & Gross, R. & Marinis, T. 2002. The processing of ambiguous sentences by first and second language learners of English. *Essex Research Reports in Linguistics* 40, 1-38.

Fernandez, E. M. 2003. *Bilingual sentence processing. Relative clause attachment in English and Spanish.* Philadelphia: John Benjamins Publishing.

Frazier, L. & Clifton, C. 1996. *Construal.* Cambridge, MA: MIT Press.

Frazier, L. & Fodor, J. D. 1978. The sausage machine: A new two-stage parsing model. *Cognition* 6, 291-235.

Gibson, E. & Schütze, C. T. 1999. Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory and Languag* 40, 263-279.

Gibson, E. & Pearlmutter, N. & Canseco-Gonzalez, E. & Hickock, G. 1996. Recency preferences in the human sentence processing mechanism. *Cognition* 59, 23-59.

Gilboy, E. & Sopena, J. M. & Clifton, C. & Frazier, L. 1995. Argument structure and association preferences in Spanish and English complex NPs. *Cognition* 54, 131-167.

Göz, İ. 2003. *Yazılı Türkçe'nin kelime sıklığı sözlüğü.* Ankara: Türk Dil Kurumu.

Hankamer, J. & Knecht, L. 1976. The role of the subject/non-subject distinction in determining the choice of relative clause participle in Turkish. In: Hankamer, J. & Assen, J. (eds.) *Harvard Studies in Syntax and Semantics.* Cambridge, MS: Harvard University. 197-219.

Kamide, Y. & Mitchell, D. 1997. Relative clause attachment: non-determinism in Japanese parsing. *Journal of Psycholinguistic Research* 26(2), 246-254.

Kimball, J. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2, 15-47.

Mitchell, D. C. & Brysbaert, M. 1998. Challenges to recent theories of crosslinguistic variation in parsing: Evidence from Dutch. In: Hillert, D. (ed.) *Syntax and semantics: A crosslinguistic perspective.* San Diego, CA: Academic Press. 313-335.

Mitchell, D. & Cuetos, F. & Corley, M. B. & Brysbaert, M. 1995. Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research* 24, 469-488.

Papadopoulou, D. & Clahsen, H. 2002. Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Essex Research Reports in Linguistics* 3, 61-92.

Roberts, L. 2003. *Second language sentence processing: the processing of relative clause attachment ambiguities and long-distance wh-dependencies by adult L2 learners of English.* [PhD dissertation, University of Essex.]

Underhill, R. 1974. Turkish participles. *Linguistic Inquiry* 5, 87-99.