Contextually Appropriate Reference Generation

ÖZGÜR YÜKSEL

Computer Engineering
Middle East Technical University, Ankara 06531, Turkey

CEM BOZSAHIN

Computer Engineering and Cognitive Science
Middle East Technical University, Ankara 06531, Turkey

(*Received 16 November2001*)

Abstract

We describe a system for contextually appropriate anaphor and pronoun generation for Turkish. It uses Binding Theory and Centering Theory to model local and nonlocal reference. We describe the rules for Turkish, and their computational treatment. A cascaded method for anaphor and pronoun generation is proposed for handling pro-drop and discourse constraints on pronominalization. The system has been tested as a standalone nominal expression generator and also as a reference planning component of a transfer-based MT system.

1 Introduction

In the process of natural language generation, reference planning is a crucial phase after utterance planning for producing an utterance in a contextually appropriate form. Utterance planning may involve goal and subgoal selection (McKeown 1985; Derr and McKeown 1984), ordering (Hovy 1988), and determining the information structure (e.g., topic and focus). Reference planning aims to preserve discourse salience of entities in a planned discourse and natural flow of the discourse by choosing forms of reference, e.g., overt and zero pro-forms, and referring expressions.

In this study, we assume Chomsky's classification of pro-forms; in Binding Theory (Chomsky 1981), reflexives and reciprocals are considered anaphors, distinct from other pronominals. Our aim is to generate anaphors and pronouns in this sense to provide a natural flow to a planned Turkish discourse.

Turkish is a pro-drop language; pronominal subjects, objects, and specifiers of possessive NPs can be dropped. Subject pro-drop is locally identifiable by agreement

[0] * Thanks to Ümit Turan, Massimo Zancanaro and an anonymous referee for comments and suggestions; and to Dilek Hakkani and Kemal Oflazer for material related to this research. It was supported in part by NATO Science Division SfS III (contract: TU-LANGUAGE). Yüksel's current address: Oracle Ltd., Ugur Mumcu Cad. 88, Ankara 06700, Turkey.

morphology on the verb, but definite and indefinite null objects are not recoverable from clause level marking. Pro-drop may be done for stylistic reasons, but in many cases, it is required in a certain grammatical or information structure. Conversely, the information structure of an utterance may require overt realization of a pronoun. Appropriate use of pro-drop is included as part of the reference planning system.

Consider examples (1) and (2).[1] In (1), the anaphor *kendi* can only refer to *Fatma* and not to another entity in the discourse. It is bound in the local domain. On the other hand, *onu* in (2) may be bound to a nonlocal clausal argument (i) or to some other entity in the discourse (j), but not to a local argument (k).

(1) *Fatma*$_i$ *kendi-ne*$_{i/*j}$ *güven-ir*
F.NOM herself-DAT trust-AOR
'Fatma trusts herself.'

(2) *Ayşe*$_j$ *kenar-da* *otur-uyor-du*
A.NOM side-LOC sit-PROG-PAST
'Ayşe was sitting by the corner.'
Murat$_i$ *koş-ar-ken* *Fatma*$_k$ *o-nu*$_{i/j/*k}$ *izl-iyor-du*
M.NOM run-AOR-CONV F.NOM he-ACC watch-PROG-PAST
'While Murat was running, Fatma was watching him/her.'

As for pro-drop, consider discourse (3). Pro-drop avoids reemphasizing the salient entity in the discourse. In this example, the topic is always *Murat*. The repetition in subsequent utterances is very unnatural. The use of overt pronouns is just as bad (4).

(3) *Okul-a* *gid-iyor-du* *Murat*$_i$
school-DAT go-PROG-PAST M.NOM
'Murat was going to school.'
Murat$_i$ *okul-u* *sev-iyor-du*
M.NOM school-ACC like-PROG-PAST
'Murat liked the school.'
Murat$_i$ *geç* *kal-mış-tı*
M.NOM late stay-AUX-PAST
'Murat has been late.'
Murat$_i$ *acele* *et-meli-ydi*
M.NOM hurry make-NEC-PAST
'Murat had to hurry.'

(4) *Okul-a* *gid-iyor-du* *Murat*$_i$
School-DAT go-PROG-PAST M.NOM
'Murat was going to school.'

[1] We use the following abbreviations in glosses. ACC: accusative, AGR: agreement, AOR: aorist, AUX: auxiliary, CAUS: causative, COM: comitative, CONV: converb, DAT: dative, GEN: genitive, INF: infinitive, LOC: locative, NEG: negative, NEC: necessitative, NOM: nominative, PERS: person, PERS3p: third person plural, PLU: plural, POSS: possessive, PROG: progressive, RECIP: reciprocal, REFLEX: reflexive, REL: relative, SUB: subordinate marker.

O_i *okul-u* *sev-iyor-du*
He school-ACC like-PROG-PAST
'He liked the school.'
O_i *geç* *kal-mış-tı*
He late stay-AUX-PAST
'He has been late.'
O_i *acele* *et-meli-ydi*
He hurry make-NEC-PAST
'He had to hurry.'

Despite the felicitious English glosses, discourse (4) is ungrammatical; use of overt pronouns in Turkish signals the introduction of (or shifting to) a new focus, topic (Enç 1986), or beginning of a new discourse segment (Turan 1995). If the information structure is retained in the subsequent utterance, zero pronouns (denoted by ϕ in the examples) are used instead. Pronouns in (4) attempt to emphasize what is already set as the topic. The natural way to express this discourse is (5):

(5) *Okul-a* *gid-iyor-du* *Murat$_i$*
School-DAT go-PROG-PAST M.NOM
'Murat was going to school.'
ϕ_i *Okul-u* *sev-iyor-du*
School-ACC like-PROG-PAST
'(He) liked the school.'
ϕ_i *Geç* *kal-mış-tı*
late stay-AUX-PAST
'(He) has been late.'
ϕ_i *Acele* *et-meli-ydi*
hurry make-NEC-PAST
'(He) had to hurry.'

The goal of the system described here is to retain, drop or replace the full NPs in generation so that the resulting discourse is quite natural. We assume that the entities of the planned discourse are decided beforehand, and that the reference planner receives them as full NPs. The system is conceived as part of the generation phase of a machine translation system based on structural transfer. The architecture of the MT system is shown in Figure 1. It translates from English (and German) to Turkish. The target text for translation is technical manuals.

The reference planner is designed to be multipurpose. In practice, utterance planning for transfer-based MT may be kept to a minimum by preserving the source language parameters for grammatical voice, topic and focus in the target language, but a text generation system will have a more sophisticated utterance planner, synthesizing most of the information from scratch. The reference planner expects the same information from the utterance planner, be they derived from the source language or from an interlingual representation such as that of Dorr (1993), Nirenburg et al. (1992), or McKeown (1985). For this reason, literary works in Turkish have been used as control data in addition to the targeted translation domain.

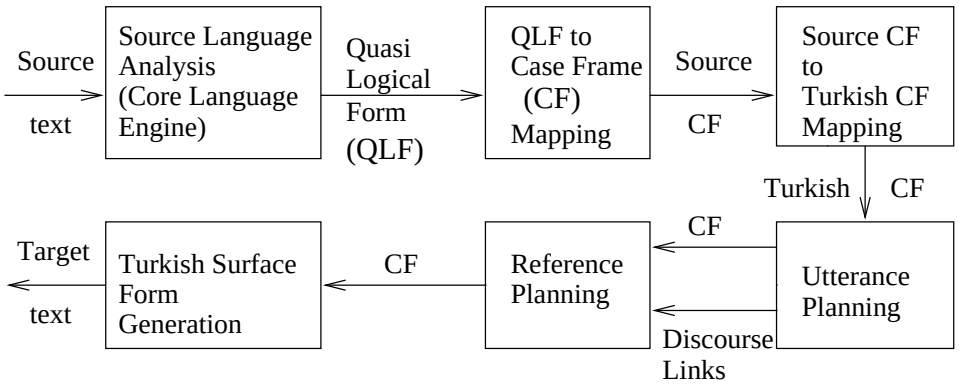

Fig. 1. English to Turkish Structural Transfer.

The reference planner consists of a set of rules for binding relations and a set of constraints on realization of centers of information in the discourse. Local reference is planned by binding rules, and nonlocal reference by an interaction of binding and centering rules. Pro-drop is handled at the final stage of reference planning. The system does not deal with cataphors, deictic pronouns, and anaphoric references like (6) which are not realized as pronouns. Table 1 shows the coverage.

(6) *Murat$_i$* *cam-ı* *kır-dı*
M.NOM glass-ACC break-PAST
'Murat broke the window.'
Cam-ı *kır-an$_i$* *düşün-sün*
glass-ACC break-REL think-IMP
'Let the one who broke it worry.'

The rest of the paper is organized as follows. Section 2 explains our use of the concepts from Binding Theory and Centering Theory. Section 3 shows how the features of Binding Theory and Centering Theory help synthesize the rules of reference generation for Turkish. Section 4 describes the operation of the reference planner. Section 5 evaluates its performance.

2 Local and Nonlocal Reference

Binding Theory of Chomsky (1981; 1986) is concerned with the interpretation of anaphors, pronouns, and referring expressions. In particular, the theory proposes three conditions as universal constraints on the interpretation of reference.

Condition A: An anaphor must be bound in its governing category (local domain).

Condition B: A pronoun must be free in its governing category. In other words, a pronoun must not be locally bound.

Table 1. *Coverage of Turkish Anaphors and Pronouns.*

Anaphors				Pronouns	
Reciprocals		Reflexives			
birbirleri	each other	kendi	self	o	he/she/it
-Iş[a]		kendi-ni	self-ACC	onu	him/her/it
		kendi-ne	self-DAT	onun	his/her/its
		kendi-nde	self-LOC	onlar	they
		kendi-nden	self-ABL	onları	them
		ϕ		onların	their
				ϕ	

[a] This is the suffix indicating a morphological reciprocal where 'I' denotes an underspecified high vowel. Its realization as ı/i/u/ü is determined by vowel harmony.

Condition C: A referring expression must be free everywhere.

Governing categories are labelled binding domains; they indicate the scope of reference relations. They can be verbal or nominal domains. In our case, conditions A, B, C are reinterpreted from the perspective of generation so that they act as reference filters on coindexation coming out of the previous phases of planning.

Centering Theory has been developed over the years by Grosz, Joshi and Weinstein (1983; 1995). It is an attempt to relate focus of attention, choice of reference, and perceived coherence of discourse. Centers of discourse are semantic entities that serve to link an utterance to other utterances. One of the tenets of the theory is the explicit correspondence between discourse salience, prominence and reference, formulated in terms of the centers in the discourse. From that perspective, centering can be seen as a theory that can be put to use to explore the coreference possibilities that are left open by Condition B and partly by Condition C: the discourse based constraints on coreference. It attempts to resolve anaphora (in its most general sense) locally within a discourse segment. For the purpose of generation, centers can be thought of as information that helps to convey the intentional state of the speaker to the hearer and to reflect her attentional state. Some aspects of the theory are summarized below. Centers of the discourse and constraints on the flow of centers play a crucial role in the theory. It makes use of the following center definitions:

$C_f(U_n)$: The list of potential centers in the utterance U_n (forward looking centers).
$C_p(U_n)$: The center selected from the C_f-list in U_n (the preferred center).
$C_b(U_n)$: The center which provides the link to the previous utterance (the backward looking center).

The following conditions hold among centers:

$C_p(U_n) \in C_f(U_n)$
$C_b(U_n) \in C_f(U_n)$
$C_b(U_n) = C_p(U_{n-1})$

Centering Theory assumes that there is a unique C_b for each utterance, and that

it only depends on the current and previous utterance. The center of the discourse may shift along the discourse segment. Since a center can be any semantic object in an utterance, there is a competition between potential centers, i.e., the elements of $C_f(U_n)$. This competition is solved by ranking the $C_f(U_n)$. In (Grosz *et al.* 1995, p.214) the ranking is SUBJECT > OBJECT(S) > OTHER. Turan (1995, p.101) proposed a hierarchy based on thematic roles: AGENT > EXPERIENCER > (Inalienable) POSSESSOR > THEME.

The relationships between the centers in consecutive utterances are called *transitions*. There are four kinds of transitions, listed below in the order of preferred ranking:[2]

Continue: $C_b(U_n) = C_b(U_{n-1})$, and this entity is the most highly ranked element of $C_f(U_n)$, i.e., $C_b(U_n) = C_p(U_n)$.
Retain: $C_b(U_n) = C_b(U_{n-1})$, but this entity is not the most highly ranked element of $C_f(U_n)$, i.e., $C_b(U_n) \neq C_p(U_n)$.
Smooth-Shift: $C_b(U_n) \neq C_b(U_{n-1})$ and $C_b(U_n) = C_p(U_n)$.
Rough-Shift: $C_b(U_n) \neq C_b(U_{n-1})$ and $C_b(U_n) \neq C_p(U_n)$.

3 Reference Planning

The reference planner we designed is a rule based system. A serial approach is suitable for this task since planning is done in phases; the rules refer to information such as governing categories and centers, which are the output of previous phases. This section describes the rules we compiled for Turkish, along with the data that motivated the formulation of the rules. For more on linguistic aspects, the reader is referred to (Turan 1995; Enç 1986; Erguvanlı-Taylan 1986).

3.1 Reciprocals

Reciprocals are assumed to be identified in the semantic representation. In our system, they are marked RECIP, and a complex NP is formed for the subject and the direct object. In surface form, the periphrastic construction (*birbirleri* 'each other') or verbal inflection marks the verb as reciprocal.[3] (7a–b) contrast the nonreciprocal and the periphrastic reciprocal.

(7) a. *Adam*$_i$ *adam-*$\imath_{j\neq i}$ *incele-di*
man man-ACC examine-PAST
'The man$_i$ examined the man$_j$.'

b. *Adam-lar* *birbirleri-ni* *incele-di-ler*
man-PLU each other-ACC examine-PAST-PERS
'The men examined each other.'

[2] The finer division of shift was proposed by Brennan, Friedman, and Pollard (1987).
[3] periphrastic forms use an auxiliary word to show a grammatical process, as opposed to a verbal inflection for the same process.

RULE RCP1: Among the argument NPs of a reciprocal verb, the object NP is realized periphrastically if the verb cannot receive morphological reciprocal.

Reciprocal can be marked as an inflection as well; the suffix -Iş is attached to the verbal stem. Verbs are marked lexically whether they can receive the morphological reciprocal. For instance, *bak* ('look at', which takes a dative object), can be inflected; its reciprocal can be realized either syntactically or morphologically (8a–b). *İncele* ('examine') can only have the periphrastic reciprocal.

(8) a. *Adam-lar$_i$* *birbirleri-ne* *bak-tı-lar*
man-PLU each other-ACC look-PAST-PERS
'The men looked at each other.'

b. *Adam-lar* *bak-ış-tı-lar*
man-PLU look-RECIP-PAST-PERS3p
'The men looked at each other'

c. *Adam-la* *kadın* *bak-ış-tı-lar*
Man-COM woman look-RECIP-PAST-PERS
'The man and woman looked at each other'

d. **Adam* *kadın-a* *bak-ış-tı-lar*
woman-DAT

RULE RCP2: If the verb is eligible for morphological reciprocal, the object NP can be dropped, yielding a complex NP as surface subject (cf. (8c) and (8d)).

3.2 Reflexives

Reflexives are assumed to be marked by coindexation in the semantic representation. We make use of the Binding Theory for reflexives. Condition A of the theory (that an anaphor must be locally bound) is interpreted as: Replace the NP that has the coindexed antecedent in its governing category by the reflexive pronoun (*kendi* 'self') along with the corresponding case marker (cf. Table 1). (9a) is realized as (9b), not (9c).

(9) a. *Murat$_j$* [$_{GC}$ *Fatma'nın$_i$* *Fatma'ya$_i$* *güvendiğini*] *bil-iyor*
M.NOM F-GEN F-DAT trust know-PROG

b. *Murat$_j$* [$_{GC}$ *Fatma'nın$_i$* *kendi-ne$_i$* *güvendiğini*] *biliyor*
REFLEX-DAT
'Murat knows that Fatma trusts herself.'

c. **Murat$_j$* [$_{GC}$ *kendinin$_i$* *Fatma'ya$_i$* *güvendiğini*] *biliyor*

RULE RFL1: If two NPs are coindexed, the succeeding NP is realized as an overt reflexive pronoun if the preceding NP is in the governing category.[4]

[4] This rule is expressed in Binding Theory in terms of the c-command relation between the binder and the bindee. For efficiency, it is possible to exploit the ranked representation and nonassociativity of the case frame coming out of utterance planning (as in our system) to implement the rule without reference to c-command.

There are cases where the reflexive is a zero anaphor (Erguvanlı-Taylan 1986). Moreover, if the reflexive in possessor position is stressed, it cannot be dropped. (10b) and (11b) are felicitous only if the reflexive pronoun is the focus.

(10) a. $[_{GC}$ *Murat*$_i$ ϕ_i *gid-eceği yer-i*] *bil-iyor*
M.NOM go-REL place-ACC know-PROG
'Murat knows where (he) will go.'

b. $[_{GC}$ *Murat*$_i$ *kendi-nin*$_i$ *gideceği yeri*] *biliyor*
self-GEN

(11) a. $[_{GC}$ *Murat*$_i$ ϕ_i *kitab-ı-nı* *oku-du*]
M.NOM book-POSS-ACC read-PAST
'Murat read his book'

b. $[_{GC}$ *Murat*$_i$ *kendi*$_i$ *kitabını okudu*]
self

RULE RFL2: If NP to be replaced by a reflexive pronoun is the possessor of a genitive construction and the genitive construction is not the focus, then zero representation is preferred if there is only one potential antecedent for the zero pronoun.

3.3 Pronouns

Binding Theory can be used to make predictions about nonlocal reference, subject to discourse conditioned exceptions. Condition B of the theory (that a pronoun cannot be locally bound) allows a flexible interpretation: Pronominalize the NP that has the coindexed antecedent out of the NP's governing category. Accordingly, (12a) must be realized as (12b).

(12) a. *Murat*$_i$ $[_{GC}$ *Fatma'ya*$_j$ *Murat'ı*$_i$ *görmesini*] *söyledi*

b. *Murat*$_i$ *Fatma'ya*$_j$ *onu*$_i$ *gör-me-si-ni* *söyle-di*
M.NOM F-DAT he-ACC see-INF-AGR-ACC say-PAST
'Murat told Fatma to see him.'

RULE PRO1: If two NPs are coindexed, the succeeding NP is realized as an overt pronoun if the antecedent is out of the governing category.

An NP must be referential if it is to be referred to by a pronoun. *turist* is nonreferential in (13); a zero pronoun is used, instead of the coreferential *onlar*.

(13) *Heryerde* *turist*$_i$ *vardır.* *Ama* *burada* ϕ_i */*onlar*$_i$ *yok*
Everywhere tourist exists . But here they not exist
'There are tourists everywhere. But here, there is none.'

RULE PRO2: If the antecedent of a pronoun is nonreferential, an overt pronoun cannot be used. A zero pronoun is employed instead.

Pro-drop is optional in certain contexts. In (14), U_2 and $U_{2'}$ are both felicitous.

(14) U_1: *Murat$_i$* *kitab-ı$_j$* *oku-du*
M.NOM book-ACC read-PAST
'Murat read the book.'

U_2: ϕ_i *Kitab-ı$_j$* *raf-a* *kaldır-dı*
book-ACC shelf-DAT put-PAST
'(He) put the book on the shelf.'

$U_{2'}$: ϕ_i ϕ_j *Raf-a* *kaldır-dı*
shelf-DAT put-PAST
'(He) put (it) on the shelf.'

RULE PRO3: The object in U_{n-1} can be expressed as a zero pronoun in U_n.

RULE PRO4: The C_b in U_n is realized as a pronoun (Gordon *et al.* 1993). For example:

(15) U_1: *Murat$_i$* *iyi* *bir* *insan-dır*
M.NOM good a man-AUX
'Murat is a good man.'

U_2: *Herkes* *o-nu$_i$/?Murat'ı* *sev-er* *(C_b=Murat)*
everyone he-ACC/M-ACC like-AOR
'Everyone likes him.'

In (16), the use of overt pronouns for both NPs is quite unnatural, and their interpretation is ambiguous. In fact, the one which is not the C_b is dropped, cf. the discourse salient version of (16b) in (16c):

(16)a. U_1: *Ali$_i$* *Mehmet'i$_j$* *gör-dü*
A.NOM M-ACC see-PAST
'Ali saw Mehmet.'

b. U_2: ?? *O_i ona$_j$ selam ver-di* *(C_b =Mehmet)*

c. $U_{2'}$: ϕ_i *O-na$_j$* *selam* *ver-di*
he-DAT hello give-PAST
'(He) said hello to him.'

There are exceptions to PRO4. Even though there is no other overt pronoun in (17), the second utterance is still unacceptable:

(17) U_1: *Ali$_i$* *Mehmet'i$_j$* *gör-dü*
A.NOM M-ACC see-PAST
'Ali saw Mehmet.'

U_2: *O_i* *çevre-si-ne* *bakın-dı*
he.NOM surrounding-POSS-DAT look-PAST
'He looked around.'

This indicates that the problem is not only the number of NPs competing for pronominalization; the interpretation of the overt pronoun is ambiguous, that is, there is more than one potential C_b. The overt pronoun in the second utterance of (17) is dropped.

RULE PRO5: An NP can be realized as an overt pronoun only if there is no other overt pronoun that is competing for C_b status, and if it is Discourse-old in U_{n-1}.

If the topic is always the same throughout the discourse segment, i.e., if there is always a Continue-subj transition as in discourse (3) and its pronominalized version (4), the NP that is the C_b must be dropped, as in (5).

RULE PRO6: If there is a Continue-subj transition, a zero pronoun is used for the subject.

Rule PRO6 is not sufficient to cover the usage of zero pronouns for subjects. Consider U_3 in discourse (18); the overt pronoun is ambiguous in this case. U_2–U_3 in (18) is a Smooth-Shift-subj transition. Dropping the pronoun subject in U_3 eliminates the unwanted ambiguity.

(18) U_1: *Murat$_i$* *koş-uyor-du* *(C_b= Murat)*
M.NOM run-PROG-PAST
'Murat was running.'

U_2: *Meral Hanım$_j$* *o-nu$_i$* *çağır-dı* *(C_b= Murat)*
Meral Hanım.NOM he-ACC call-PAST
'Meral Hanım called him.'

U_3: *O$_{?i/j}$* *se-si-ni* *duy-ur-am-ıyor-du* *(C_b=Meral hanım)*
She voice-POSS-ACC hear-CAUS-NEG-PROG-PAST
'She could not make herself heard.'

RULE PRO7: If there is a Smooth-Shift-subj transition, a zero pronoun is used for the subject.

Pronominalization and pro-drop are further constrained by information structure. The remainder of the rules describe exceptions to generating overt and zero pronouns.

(19) U_1: *Murat$_i$* *Fadıl$_j$* *ile* *konuş-uyor-du* *(C_b=Murat)*
M.NOM F.ACC with speak-PROG-PAST
'Murat was talking to Fadıl.'

U_2: **O$_{i/j}$* *düşün-üyor-du* *(C_b = Fadıl)*
He think-PROG-PAST
'He was thinking.'

In (19:U_2), the pronoun subject is ambiguous; it can be interpreted as *Murat* or *Fadıl*. Using a null pronoun is not a solution either. A contextually more appropriate form of (19) is (20).

(20) U_1: *Murat$_i$ Fadıl$_j$ ile konuşuyordu*
U_2: *Fadıl$_j$ düşünüyordu*

RULE PRO8: A full NP subject occurs in a Shift-nonsubj if it is not $C_p(U_{n-1})$.

In (21), we see a situation related to PRO5. Contrary to PRO5, however, dropping the first pronoun does not solve the problem. This is an exception to PRO5. The difference from the context of the rule is that, in this case, a Retain transition has occurred; a full NP is used for the subject:

(21) U_1: *Murat$_i$* *Rezzan'a$_j$* *şarkı* *söyle-me-yi* *sev-er-di*
M.NOM R-DAT song say-INF-ACC like-AOR-PAST
'Murat liked singing to Rezzan.'
U_2: * *O_j/ϕ_j onun$_i$ geleceğini parlak görüyordu* *(C_b=Murat, C_p=Rezzan)*
$U_{2'}$: *Rezzan$_j$* *o-nun$_i$* *geleceğ-i-ni* *parlak* *gör-üyor-du*
R.NOM he-GEN future-POSS-ACC bright see-PROG-PAST
'Rezzan thought he has a bright future.'

RULE PRO9: Full NP subjects are used in Retain transitions.

There are cases where the pronoun is ambiguous even if it is the only C_b in U_{n-1}, as in (22). In this case, a full NP is used (23):

(22) U_1: *Murat$_i$* *Fadıl'ı$_j$* *sev-er-di* *(C_b=Murat)*
M.NOM F-ACC like-AOR-PAST
'Murat liked Fadıl.'
U_2: ? *O_j* *güzel* *konuş-ur-du*
He.NOM beautiful talk-AOR-PAST
'He was a good speaker.'

(23) U_1: *Murat$_i$ Fadıl'ı$_j$ severdi*
U_2: *Fadıl güzel konuşurdu*

RULE PRO10: If an entity is (a) in a nonsubject position in U_{n-1}, (b) not the $C_b(U_{n-1})$, and (c) not distinguishable by Ψ-features number, person and gender from the C_p in the C_f-list, it is realized as a full NP subject in U_n.

Finally, we need the pronoun counterpart of RFL2 to account for genitive constructions with nonreflexive possessors. The rule below handles cases such as (24).

RULE PRO11: If the NP to be replaced by a pronoun is the possessor of a genitive construction and the genitive construction is not the focus, then the zero form is preferred if there is only one potential antecedent for the zero pronoun.

(24) U_1: *Ali$_i$* *yol-da* *yür-üyor-du*
A.NOM street-LOC walk-PAST
'Ali walked on the street.'
U_2: *Murat* *ϕ_i/onun$_i$* *git-tiği* *yer-i* *bil-iyor-du*
M.NOM go-REL place-ACC know-PROG-PAST
'Murat knew where (he) was going.'

We make no claims about exhaustive coverage of reference generation, but these rules seem to have worked satisfactorily on the literary and technical documents we have studied. In the next section, we describe the system. There are interactions between the rules, thus a proper ordering is required in the computational treatment of anaphor and pronoun generation.

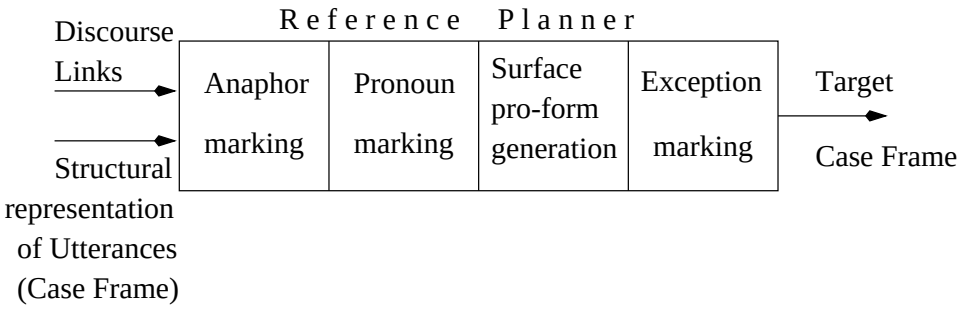

Fig. 2. Four stage Reference Planning.

4 Reference Planning System

Our goal in the design of the reference planning system was to set up a general framework of which Turkish reference planning is a particular realization. We investigated two possibilities. The first approach was deriving the dependencies between the rules from rule specifications, and compiling them into a dependency network. The rules can then be executed in an order not violating the dependencies. However, automatic detection of dependencies between the rules is a formidable task due to their reliance on various sources of information (local binding environment, lexical properties, centers, types of transitions, surface order and grammatical functions, etc.). Manual construction of the network is possible, but this would not scale up to a system with large number of rules.

We chose to evaluate the rules in a four stage process depicted in Figure 2. The first two stages decide whether an NP is to be realized as a referring expression or as a pro-form. Decisions on local reference are made first since nonlocal reference may depend on them (e.g., dependence of PRO1 on RFL1 in our case). The third stage determines whether pro-drop is to be employed for an NP that the system decided to pronominalize in the previous stages. The last stage may revert some of these decisions, e.g., constraining the drop of a pro-form (e.g., PRO5) or retracting the pronominalization of a full NP (e.g., PRO8, PRO9, PRO10).

The structural representation of utterances is given in Figure 3. The NP structure of this representation is depicted in Figure 4. The path REFERENT | CONTROL | INDEX provides local coindexation. DROP is a binary feature whose value is determined by the reference planner. The replacement of full NPs with anaphors and pronouns is not performed in place because overt and zero pronouns can be retracted by exception rules. A CONTROL structure is associated with all nominal expressions whose realization is to be decided by the reference planner (cf. POSSESSOR|CONTROL in Figure 4 for pro-drop of possessive NPs).

The discourse links required by the planner are the backward looking centers and the preferred centers. The input to the reference planner is sets of discourse segments in which every utterance is represented by the triplet $[C_f(U_i), C_b(U_i), C_p(U_i)]$. The output is the Turkish case frames in which CONTROL structures are marked with reference information. Lexical substitution for periphrastic constructions and overt

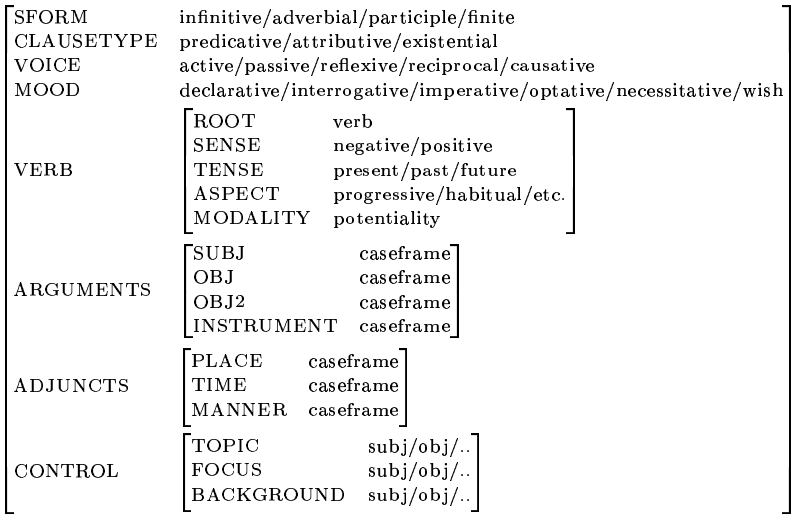

Fig. 3. The Turkish Case Frame (CF_T)

pronouns are also performed by the planner. The output is passed on to a surface form generator (Hakkani and Oflazer 1998).

Conditions on discourse appropriateness of reference are generally formulated from the perspective of the hearer (analysis). To turn these rules into computationally interpretable processes of generation, they need to be recast in a different form. There are three kinds of rules:

Overt Realization Rule A full NP is to be realized as an overt anaphor or pronoun under some specified circumstances. These rules belong to the first two stages of reference planning.

Drop Rule A full NP is to be realized as a zero anaphor or pronoun under some specified circumstances. This is relevant for pro-drop languages such as Turkish and Spanish. These rules belong to the third stage of reference planning.

Exception Rule An NP cannot be realized as a zero form, or an NP cannot be realized as a pro-form instead of a full NP, under some specified circumstances. These rules belong to the final stage of planning.

Below, we present the programming of the following rule as an example:

(25) The C_b must be realized as a pronoun rather than a full NP (Turan 1995, p.85)

Example (25) is an overt realization rule. We can express it as a generation rule:

(26) If the C_b in U_n is a full NP, pronominalize it.

The rules are converted into condition-action pairs. Roughly, (26) is coded as:

(27) **If** is_full$(C_b(U_n))\wedge$ is_NP$(C_b(U_n))$ **then** mark$(C_b(U_n), PRO)$.

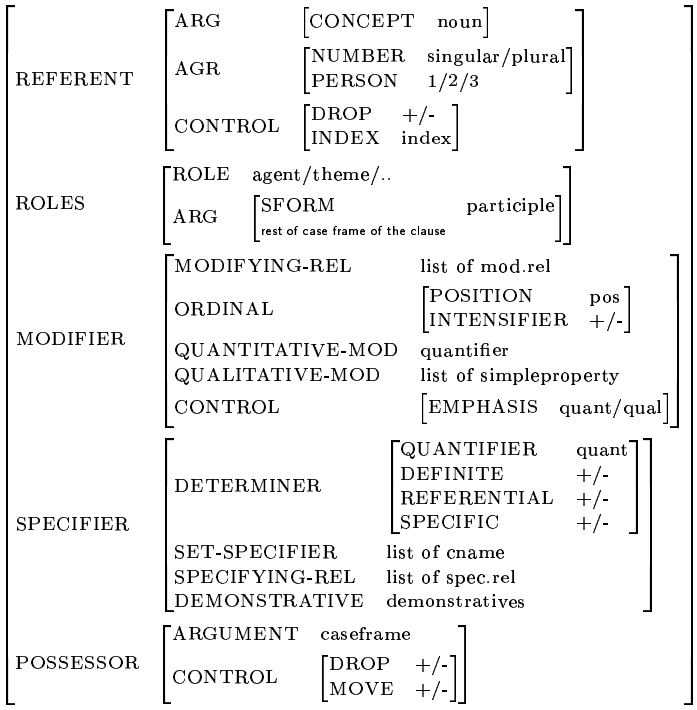

Fig. 4. Turkish Case Frame (CF_T) for NPs

Some rules have options. Consider (28). Both full NP and the dropped versions are appropriate in the discourse. The corresponding rule (PRO3) has switches to generate only full NPs or zero/overt pronouns on demand.

(28) U_1: *Fatma$_i$* *top-u$_j$* *fırla-t-tı*
F.NOM ball-ACC bolt-CAUS-PAST
'Fatma threw the ball.'

U_2: *Murat* *top-u$_j$/ϕ_j* *yakala-dı*
M.NOM ball-ACC catch-PAST
'Murat caught (the ball).'

Handling exceptions in a separate stage can be justified as follows. Suppose $R = \{rule_1, \ldots, rule_n\}$ is the set of n rules and $C_k = \{cond_{k1}, \ldots, cond_{km}\}$ is the set of m conditions for $rule_k$ in the form

$$rule_k : \textbf{if} \quad C_k \quad \textbf{then} \quad action_k$$

Let $E = \{exception_1, \ldots, exception_p\}$ be the set of exceptions for all the rules. Suppose that exceptions are not kept separately but integrated into the rules. Then, each rule would have the form:

$$rule_k : \textbf{if} \quad C_k \wedge \neg E' \quad \textbf{then} \quad action_k \text{ where } E' \in powerset(E)$$

which results in repetition of some exceptions without any real work being done. Cascaded treatment of nonmonotonic decisions simplify the rules as well; they need only to refer to their relevant properties. Therefore, we prefer to separate the pro-

cessing of exceptions. When we handle exceptions separately, the whole exception set must be applied on the results from previous stages because an exception may not be specific to a rule.

5 Evaluation

For evaluation, we conducted three different kinds of experiments. We used the data in two different modes. In experiment 1, we looked at four Turkish novels. Literary text provided a performance measure which is independent of the MT system. It served as the control data. The performance of the system is compared with the anaphor and pronoun usage in the original text. In experiment 2, we did our own evaluation of the output for the targeted translation domain (computer manuals). The technical text was presented to the system independent of its source language parameters.[5] The results are classified according to the acceptability and naturalness of the output. In experiment 3, we tested subject agreement on system's performance on the technical text. In subject evaluation, one group of subjects rated the system's output (experiment 3.1), and two other groups were asked to do the task performed by the system (experiment 3.2). Experiment 3.1 gives a measure of appropriateness for the system's output. Experiment 3.2 is an indicator for the degree to which subjects would do the same task in the same manner as our system.

Literary examples include several stylistic variations that are not covered by the system, therefore the results are classified as 'unacceptable', 'acceptable but not preferred in original text', and 'acceptable and preferred in original text'. Timing is also critical in the system since it is one of a series of modules in the planning component. Table 2 shows the evaluation times for fifteen randomly selected discourse segments from the literary text.[6]

All examples make extensive use of anaphors and pronouns. 7.4 per cent of the pro-forms were anaphors, and 92.6 per cent were pronouns in experiment 1 (approximately 10 per cent and 90 per cent for experiments 2 and 3). To prepare the input set for experiments 1 and 2, all overt and zero pro-forms were replaced by the corresponding nominal expressions. In Table 3 and Table 4, we show the results of experiments 1 and 2 for utterances subsequent to the first utterance in the discourse segment. Discourse links provide no clues for the first utterance in the discourse, and only local reference is resolved. Some single utterance discourses are chosen to indicate the performance in this respect. All unreported first utterances in multi-utterance discourses gave acceptable results. We did not take them into account in performance evaluation.

Table 3 lists the evaluation of some utterances in sample literary discourses. The first column is the example number (E_i) and the second column is the utterance

[5] In the version we intend to use for the MT system, source language parameters are preserved (except the pronominalizations) so as to simplify utterance planning and derivation of centering information.

[6] The experiments were done on Sparcstation 4 running SICStus Prolog.

Table 2. *Evaluation Times for Sample Discourses.*

Discourse Number	Number of Utterances	Evaluation Time (msec)	Average time per Utterance (msec)
1	1	022	22.00
2	2	132	66.00
3	2	066	33.00
4	2	104	72.00
5	2	091	45.50
6	2	107	53.50
7	4	277	69.25
8	2	125	62.50
9	2	097	48.50
10	3	145	48.33
11	2	076	38.00
12	2	112	56.00
13	4	230	57.50
14	4	316	79.00
15	3	160	53.33

number in the corresponding example ($E_{i,j}$). Subsequent columns show our evaluation of the output, and the last column shows the match with the novelist's preference. Among the unacceptable utterances, the majority is grammatical but quite marked in the current context. We considered them unacceptable because we are concerned with discourse level felicity. In Table 3, only $E_{2,1}$ is certainly nonfelicitous. $E_{4,2}$ is U_2 in (29), and it is not preferable. The preferred version is $U_{2'}$.

(29) U_1: *Cevdet Bey* *Nigan'ı* *sev-eceği-ni* *sez-iyor-du* (Pamuk, 1983)
N-ACC love-SUB.AGR-ACC feel-PROG-PAST
'Cevdet Bey had the feeling that he would love Nigan.'

U_2: *Nigan'ı* *sev-mek* *iste-diği-ni* *düşün-müş-tü*
N-ACC love-INF want-SUB.AGR-ACC think-AUX-PAST
'He thought he wants to love Nigan.'

$U_{2'}$: *O-nu* *sev-mek* *iste-diği-ni* *düşün-müş-tü*
She-ACC love-INF want-SUB.AGR-ACC think-AUX-PAST
'He thought he wants to love her.'

For example, $E_{13,4}$ is U_4 in (30). Although it is acceptable and preferable, $U_{4'}$ is used in the original text. This is a stylistic variation. Table 4 shows the success rates for experiment 2. The results are classified as acceptable or unacceptable because we observed no stylistic variation in the text.

(30) U_1: *Isa* *Uğur'a* *düşkün-dü* (U. Ersoy, 1983)
U-DAT devoted-AUX
'Isa was devoted to Uğur.'

U_2: ϕ *Uğur'u* *sev-er-di*
U-ACC love-AUX-PAST
'(He) liked Uğur.'

Table 3. *Results of Experiment 1 (Literary Text Experiment).*

E_i	$E_{i,j}$	Unacceptable	Acceptable but not Preferred in Text	Acceptable and Preferred in Text
1	1			√
2	1	√		
2	2	√		
3	2		√	
4	2	√		
5	2	√		
6	2			√
7	2	√		
7	3		√	
7	4		√	
8	2			√
9	2			√
10	2		√	
10	3			√
11	2			√
12	2			√
13	2			√
13	3			√
13	4		√	
14	2	√		
14	3	√		
14	4	√		
15	2			√
15	3		√	

U_3: *Uğur da o-na tap-ar-dı*
also he-DAT adore-AOR-PAST
'Uğur liked him too.'

U_4: *Isa o-nun kahraman-ı-ydı*
he-GEN hero-POSS-AUX
'Isa was his hero.'

$U_{4'}$: *O Uğur'un kahraman-ı-ydı*
He U-GEN hero-POSS-AUX
'He was Uğur's hero.'

In experiment 3, we asked 43 native speakers 125 reference cases in 106 multi-utterance discourse segments taken from the manually translated technical text. In experiment 3.1, 16 subjects were asked to rate the system's performance on the text. System's choices of reference were highlighted in the questionnaire to signify what is being tested, but in order to eliminate bias, the subjects were not told that these references were computer generated. This experiment provides a basis for the weaker claim that the native speakers find the system's output contextually appropriate. The result of this test is shown in Table 5. Each count in the table indicates the number of times the subject agreed with the system out of 125 cases. The average agreement is 72.55 per cent with standard deviation $s_{n-1} = 13.29$. From these figures, the 95 per cent confidence interval for expected agreement is

Table 4. *Results of Experiment 2 (Technical Text Experiment).*

E_i	$E_{i,j}$	Unacceptable	Acceptable
1	1		√
2	1	√	
3	1	√	
4	2	√	
5	2		√
6	2		√
7	2		√
8	2		√
9	1		√
10	1		√

Table 5. *Results of Experiment 3.1 (Subject Rating of the System's Output).*

Subject	Total Agreement	Total Disagreement	Percent Agreement
1	85	40	68
2	92	35	73.6
3	83	42	66.4
4	65	60	52
5	99	26	79.2
6	114	11	91.2
7	85	40	68
8	69	56	55.2
9	103	22	82.4
10	102	23	81.6
11	66	59	52.8
12	79	46	63.2
13	95	30	76
14	109	16	87.2
15	85	40	68
16	120	5	96

65.47–79.63 per cent, i.e., this interval will cover the expected percent agreement for the entire population of such cases with a probability of 0.95.

A stronger claim could be made, if, not only the native speakers find the output appropriate, they would use the same reference strategy as the system. This way, one can argue that the system's performance is essentially indistinguishable from that of native speakers. A precondition for this claim is that the subjects have a reasonable degree of agreement among themselves about discourse appropriate use of referring expressions, pronouns, pro-drop and overt anaphora. In order to test this claim, we conducted experiment 3.2 with 27 subjects. The subjects were divided into two groups, and each group were given a different set of 53 discourses. The references in the discourse were left blank, and each subject was asked whether the referring expression, zero or overt anaphor, or a pronoun provided in the options is suitable in the discourse. We checked their agreement with the Kappa test.

Table 6. *Overall Classification of Results.*

	Experiment 1	Experiment 2	Experiment 3.1
Unacceptable	33.3%	30.0%	27.45%
Acceptable but not preferred	25.0%		
Acceptable and preferred	41.7%	70.0%	
Acceptable	66.7%	70.0%	72.55%

The Kappa statistic is increasingly used in NLP studies for testing subject agreement, e.g. (Carletta 1996; Walker *et al.* 1997; Yeh and Mellish 1997). It is a measure of agreement on nominal categories which takes precautions against agreement by chance (Siegel and Castellan 1988). The kappa values for the two subject groups in experiment 3.2 were $K_1 = .28$ and $K_2 = .372$ with variances $s_1^2 = .000134$ and $s_2^2 = .00013$. These kappa values indicate low agreement among subjects. The kappa significance statistics were $z_1 = 24.08$ and $z_2 = 32.69$, which imply that although the agreement is low, it is significantly greater than zero after corrections for chance agreement. But low agreement prevents us from pursuing the stronger claim. At closer inspection, one sees "near misses" among subjects; one subject drops a pronoun where another voted for overt realization, indicating optionality of some pro-drop phenomena. Some subjects overrode the reflexive option *kendi* but used the closely related nonreflexive *kendisi*, which is a way of pronominal reference in polite or respectful speech. Furthermore, some subjects filled in a deictic alternative not provided in the questionairre as an option, such as 'this book' instead of 'the book'. Since our system offered no treatment of deixis, we felt it necessary to look at stronger claims only after more integrative approaches to discourse realization incorporating nominal anaphora, temporal anaphora, cue phrases, and deixis are developed. Lack of extensive corpus studies on Turkish discourse is also an impediment in this regard.

6 Conclusion

Contextually appropriate generation of pronouns and anaphors can be modeled as a combination of interpreting the local and nonlocal cues on planned grammatical and information structure. We showed a system of rules for Turkish based on Binding Theory and Centering Theory. Formulating discourse salience and grammatical constraints on pro-forms in these frameworks provides a good (and extendible) methodology for computational treatment of reference planning. Using them 'in reverse' for generation acts as a reference filter constraining the production of referring expressions and pronouns. Two stage decisions, such as pronominalization then dropping, and conflicting demands on the pronominalizations require a cascaded treatment of reference generation.

The system has been tested as part of an MT system, and also as a standalone generator of pro-forms. Table 6 summarizes the results of the experiments. Average response time was short enough to include the system as an online filter in

the generation phase of the MT system. Approximately 70 per cent success rate is obtained in generating natural discourses. Similar results have been reported by Yeh and Mellish (1997). This is an acceptable rate for an MT system with a post-editing component, and it is significant that the exact match in the literary text ranked highest. Although the subject tests indicate that the output is considered appropriate, it is also clear that native speakers use other sources of information in reference planning as well. We believe that adding more rules which make use of richer information sources, rather than adding more constraints on the rules, is a better alternative in a system of reference generation to achieve higher levels of success. These sources can be (i) linguistic, e.g., lexical relations between non-nominal entities, or referential forms of other arguments with different grammatical functions, (ii) nonlinguistic, e.g., a mental model of the nonlinguistic context for deictic reference, or (iii) mixed, as in rhetorical structure of discourse (Mann and Thompson 1987), which can handle discourse links in nonadjacent utterances (Not and Zancanaro 1996).

References

Brennan, S. E., Friedman, M.W. and Pollard, C.J. 1987. A Centering Approach to Pronouns. In: *Proceedings of 25th Annual Meeting of ACL*, Stanford.

Carletta, J. 1996. Assessing Agreement on Classification Tasks: The Kappa Statistic. *Computational Linguistics* **22**(2):249–254.

Chomsky, N. 1981. *Lectures on Government and Binding.* Dordrecht: Foris.

Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origin and Use.* Cambridge, MA: MIT Press.

Derr, M. A. and McKeown, K.R. 1984. Using Focus to Generate Complex and Simplex Sentences. In *Proceedings of the 10th International Conference on Computational Linguistics*, Stanford.

Dorr, B. J. 1993. *Machine Translation: A View from the Lexicon.* Cambridge, MA: MIT Press.

Enç, M. 1986. Topic Switching and Pronominal Subjects in Turkish. In D. Slobin and K. Zimmer (eds.): *Studies in Turkish Linguistics.* Amsterdam: John Benjamins.

Erguvanlı-Taylan, E. 1986. Pronominal Versus Zero Representation of Anaphora in Turkish. In D. Slobin and K. Zimmer (eds.): *Studies in Turkish Linguistics.* Amsterdam: John Benjamins.

Gordon, P. C., Grosz, B. and Gilliom, L. 1993. Pronouns, Names and the Centering of Attention in Discourse. *Cognitive Science* **17**(3):311–347.

Grosz, B. J., Joshi, A.K. and Weinstein, S. 1983. Providing a Unified Account of Definite Noun Phrases in Discourse. In *Proceedings of the 21st Annual Meeting of ACL*, Cambridge.

Grosz, B. J., Joshi, A.K. and Weinstein, S. 1995. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics* **21**(2):203–225.

Hakkani, D. Z. and Oflazer, K. 1998. Tactical Generation in a Free Constituent Order Language. *Natural Language Engineering* **4**(2).

Hovy, E. 1988. *Generating Natural Language Under Pragmatic Constraints.* Hillsdale, NJ.: Lawrence Erlbaum.

Mann, W.C. and Thompson, S.A. 1987. Rhetorical Structure Theory: A Theory of Text Organization. In Polanyi L. (ed.): *The Structure of Discourse.* Ablex.

McKeown, K. 1985. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text.* Cambridge, UK: Cambridge University Press.

Nirenburg, S., Carbonell, J., Tomita, M. and Goodman, K. 1992. *Machine Translation: A Knowledge-based Approach*. San Mateo, CA: Morgan Kaufman.

Not, E. and Zancanaro, M. 1996. Exploiting the Discourse Structure for Anaphora Generation. In *Proceedings of Discourse Anaphora and Anaphora Resolution Colloquium*, (DAARC96), Lancaster.

Siegel, S. and Castellan, N.J. 1988. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw Hill.

Turan, Ü. D. 1995. Null vs. Overt Subjects in Turkish Discourse: A Centering Analysis. Ph.D. thesis, University of Pennsylvania.

Walker, M.A., Litman, D.J., Kamm, C.A. and Abella, A. 1997. PARADISE: A Framework for Evaluating Spoken Dialogue Agents . In *Proceedings of ACL/EACL*, Madrid.

Yeh, C.-L. and Mellish, C. 1997. An Empirical Study on the Generation of Anaphora in Chinese. *Computational Linguistics* **23**(1):169–190.