Cakchiquel Reference and Centering Theory

Nancy Hedberg and Sandra Dueck

Simon Fraser University

We apply Centering Theory to a 1500-word spoken narrative by a native speaker of Cakchiquel Mayan. The purpose is to better understand discourse constraints on the form of referring expression. After an introduction to the theory, it is shown that backward-looking centers are very often encoded by zero pronouns, and that center shifts more often use full pronouns and full noun phrases than do center continuations and retains. Full noun phrase continues correlate with discourse segment openings, and full pronoun continues correlate with discourse segment closings and with transitions from full noun-phrase to zero pronoun references.

1. Introduction.

This paper brings Centering Theory to bear for the first time on reference tracking in an American language. The text is a 1500 word (230 sentence) spoken re-telling of the movie, *El Norte* by a native speaker of Cakchiquel Mayan.[1] The narrative is the story of a Mayan brother and sister, who flee from Guatemala, illegally enter the United States and try to survive there. The movie is a tragedy in that the sister dies at the end.

The language we investigate is Cakchiquel, a morphologically ergative language, with subject and objects marked for person and number on the verb. Ergative marking also occurs for possessors and arguments of relational nouns. Because of the person-number marking on verbs and nouns, speakers have a choice as to whether to express an argument of either a verb or a noun as a full noun phrase, a full pronoun, or a zero anaphor. The primary purpose of this paper is to determine discourse factors contributing to this choice.

Centering Theory is a theory first developed by computational linguists for tracking the focus of attention in discourse (Joshi and Weinstein 1981, Brennan, Friedman and Pollard 1987). Centering Theory is intended as a model of discourse coherence within discourse segments rather than between discourse segments (that is, local coherence rather than global coherence), but in this text we are interested in boundaries between segments, so we will be exploring ways that Centering Theory can be extended to account for aspects of global discourse structure.

We will first present the theory, explaining it with examples from the Cakchiquel text. Then we will turn to a discussion of our method of coding our data and to some conclusions we can draw from the data. We will conclude with a discussion of some open issues that emerged in applying the theory to the data.[2]

2. Centering Theory.

We will discuss four aspects of Centering Theory: Constraints, Ranking, Rules, and Transitions. Before beginning this discussion, we will introduce and illustrate some of the key terms.

2.1. Key Terms in Centering Theory.

The basic premise of Centering Theory is that in order to maintain local coherence in their utterances, speakers will make a link between a referent in their current utterance and a referent in their previous utterance. This referent is called "the backward-looking center" or Cb.

At the same time speakers draw on referents from their current utterance to frame their next one. Referents from the current utterance are called "forward-looking centers" or Cfs. Examples of Cbs and Cfs are shown in (1).[3]

(1) (i) ni-Ø-ki-yaqa ki-winaq
INCPL-3sA-3pE-leave 3pE-people
'They left behind their people'

Cb: [Rosa & Enrique]

Cf: [Rosa & Enrique, their people]

(ii) i x-e-el kere pa-norte

and CPL-3pA-leave there in-north

'and they left down there for the north.'

Cb: [Rosa & Enrique]

Cf: [Rosa & Enrique, Guatemala, the north]

In sentence (i), the plural agent marker on the verb refers to Rosa & Enrique from the previous utterance, i.e. the backward-looking center or Cb. The list of all the referents in the current utterance—Rosa & Enrique, their people—constitutes the most likely source of referencs for the following utterance. Hence they are the Cf's or forward-looking centers. Sentence (ii) shows a similar pattern. Rosa & Enrique continue to be the Cb, and Rosa & Enrique, Guatemala and the north constitute the list of all the referents in the current utterance or the Cf's.

It can be seen that the backward-looking centers maintain continuity, and the forward-looking centers introduce new referents as potential future centers. The first referent on the list of forward-looking centers is called the "preferred center," or Cp. This is the referent most likely to become the Cb of the following utterance. The order in which referents are ranked is determined by their grammatical relation, discussed in section 2.2 below. There is a fourth key term--the Ct or "center transition," which we'll come back to when we discuss transitions.

2.2. Constraints.

From this analysis of how centers maintain discourse continuity, the three constraints in (2) are derived.

(2) For each utterance U_i in a discourse segment D consisting of utterances $U_1,...,U_m$:

1. There is at most one backward-looking center, $Cb(U_i)$.
2. Every element of the forward-looking centers list, $Cf(U_i)$, must be realized in U_i
3. The backward-looking center, $Cb(U_i)$, is the highest-ranked element of $Cf(U_{i-1})$ that is realized in U_i.

Constraint 1 states that an utterance has at most one backward-looking center. In the first sentence of (1) the single Cb is Rosa & Enrique, and the same is the case for the second sentence in (1).

Constraint 2 states that every element of the forward-looking centers list must be realized in the utterance. In both sentences in (1), Rosa & Enrique are on the forward-looking centers list even though they are realized as zero pronouns in the utterance.

Constraint 3 entails that the Cb of the current utterance is most likely to be the preferred center or Cp of the previous utterance. In the second sentence of (1), Rosa & Enrique is the Cb, and it is also at the top of the Cf list of the first sentence, as the Cp, because it is the highest ranked element in the first sentence.

2.3. Ranking of the Forward-Looking Centers.

In order to arrive at this ranking of the elements, the set of forward-looking centers is an ordered list, as shown in (3).

(3) Ranking of the Forward-looking centers:[4]

Subject > Object(s) > Other

Subjects are ranked higher than objects, and objects are ranked higher than other relations (including arguments in subordinate clauses).

2.4. Pronoun Rule

Finally, in addition to the ranking of subjects, objects and other, centering theorists have formalized two rules for the assignment of centers. The first is called the "pronoun rule," shown in (4):[5]

(4) Pronoun Rule:

For each U_i in a discourse segment D consisting of utterances U_i,..., U_m, if some element of $\{Cf(U_{i-1})\}$ is realized as a zero pronoun in U_i, then so is $Cb(U_i)$. If there are no zero pronouns in U_i, but there is an entity realized as a pronoun in U_i, then so is $Cb(U_i)$.

This rule states that the Cb prefers zero pronouns to full pronouns, and full pronouns to full NP's. This rule captures the intuition that pronouns—and especially zeros—are used to indicate a high degree of salience in the discourse context, and the Cb is supposed to be the most salient referent expressed in an utterance. We'll come back to Rule 2—the Ordering Rule—after we've discussed transitions.

2.5. Transitions.

Based on the options available to speakers for maintaining continuity, the model identifies four types of transition: continue, retain, smooth-shift, and rough-shift. These indicate the local discourse relations between a current utterance and the preceding utterance, and they are designated Ct, or center transitions. We will illustrate these each in turn.

2.5.1. Continue

First of all, a continue transition type occurs when the Cb of the current utterance is the same as the Cb of the previous utterance, and the Cb of the current utterance is the same as the Cp of the current utterance; that is, when the Cb is realized in subject position. An example of this is shown in (5)

(5) (i) x–Ø-ra-jo-ta ria achike-ta x-Ø-u-ban che

CPL-3esA-3sE-want-IRR he what-IRR CPL-3sA-3sE-do to-3sE

He wanted to do something about it.

Cb: [Enrique]

Cf: [Enrique, something, Rosa's death]

Ct: [smooth shift]

(ii) x-Ø-oq'

CPL-3sA-cry

He cried.

Cb: [Enrique]

Cf: [Enrique]

Ct: [continue]

The Cb's are the same in both sentences--Enrique. And the Cb of sentence (ii) is the same as its Cp--Enrique. Continues occur when the speaker continues talking about the same entity that he was talking about before. Continues have been shown to be the most frequent transition type used by speakers (Hurewitz 1998: 280).

2.5.2. Retain.

A retain transition type is realized when the Cb of the current utterance is the same as the Cb of the preceding utterance, but the Cb of the current utterance is not the same as the Cp of the current utterance; that is when the Cb is not realized in subject position. An example of this is shown (6):

(6) (i) matiox che ri-dios ma-x-Ø-ki-qaj-ta-na ri-poq
thanks to def-God neg-CPL-3sA-3pE-borrow-IRR-MOD def-money
'Thanks to God they didn't have to borrow the money.

Cb: [Rosa & Enrique]
Cf: [Rosa & Enrique, money]
Ct: [continue]

(ii) sino k'o jun mismo ri ki-winaq,
since exist one same def 3pE-people

jun señora ri kiere-desir
one woman def that-is-to-say

akwi ki-ch'alal ki-banon chi-la watemal.
how 3pE-relative 3pE-doing to-there Guatemala

'since there was one of their people a woman, who that is to say was just like their relative down there in Guatemala.'

Cb: [Rosa & Enrique]
Cf: [the woman, Rosa & Enrique, Guatemala]
Ct: [retain]

Here, in sentence (ii), the Cb--Rosa & Enrique--is not equal to the Cp--the woman. The Cp introduces a new referent. As the discourse proceeds, attention shifts onto this new discourse

entity. Retains have been suggested to occur in contexts like this: when the speaker has been talking about an entity but intends to signal that he will be making a shift onto a new entity.

2.5.3. Smooth Shift.

Smooth shifts occur when the Cb of the current utterance is not the same as the Cb of the previous utterance, and when the Cb of the current utterance IS the same as the Cp of the current utterance, as in (7):

(7) (i) i x-Ø-kom ri-r-ana ri-enrik
and CPL-3sA-die def-3sE-sister def-Enrique
'And Enrique's sister died.'

Cb: [Rosa]
Cf: [Rosa, Enrique][6]
Ct: [continue]

(ii) x–Ø-ra-jo-ta ria achike-ta x-Ø-u-ban che
CPL-3sA-3sE-want-IRR he what-IRR CPL-3sA-3sE-do to-3sE
'He wanted to do something about it.'

Cb: [Enrique]
Cf: [Enrique, something, it]
Ct: [smooth shift]

Here Rosa was the Cb of sentence (i) and Enrique is the Cb of sentence (ii). Enrique appears on the list of forward looking centers in sentence (i), and he becomes the new subject of sentence (ii)

and thus the new Cp, so we have a smooth shift. Smooth shifts occur when the speaker has started talking about a new referent in such a way as to indicate that he intends to continue focusing on it.

2.5.4. Rough Shift.

Finally, we come to the fourth category of transitions: rough shifts. A rough shift occurs when the Cb of the current utterance is not the same as the Cb of the previous utterance, and when the Cb of the current utterance is not the same as the Cp of the current utterance. An example of a rough shift in our data is shown in (8):

(8) (i) kwan enrik x-Ø-u-tsu r-ana?
when Enrique CPL-3sA-3sE-see 3sE-sister

x-Ø-u-tsu xa najin kom.
CPL-3sA-3sE-see just cont die

'When Enrique saw his sister, he saw her just dying.'

Cb: [Rosa]
Cf: [Enrique, Rosa]
Ct: [retain]

(ii) i x-Ø-kom r-ana?
and CPL-3sA-die 3sE-sister
'and his sister died.'

Cb: [Enrique]
Cf: [Rosa, Enrique]

Ct: [rough shift]

Rosa was the Cb of the previous utterance--sentence (i), and Enrique is the Cb of the current utterance--sentence (ii), but Enrique is NOT the Cp of the current utterance, Rosa is. As the discourse proceeds, attention shifts onto Rosa.

Rough shifts are supposed to occur when the speaker has begun talking about a new entity (here, Enrique) but is only doing so momentarily (as here, we shift immediately onto Rosa). Rough shifts have been reported to be rare or nonexistent in natural discourse. We only have four in our data.

2.6. Summary of Transition Types.

A summary of the four transition types is shown in Table 1:

	$Cb(U_i) = Cb(U_{i-1})$ OR $Cb(U_{i-1}) = [?]$	$Cb(U_i) \neq Cb(U_{i-1})$
$Cb(U_i) = Cp(U_i)$	CONTINUE	SMOOTH-SHIFT
$Cb(U_i) \neq Cp(U_i)$	RETAIN	ROUGH-SHIFT

Table 1. Centering Transition States

2.7. The Ordering Rule.

Centering theorists rank these four transition types in terms of how strongly they maintain discourse coherence, and this brings us back to the second rule for the assignment of centers (the first was the Pronoun Rule)—the Ordering Rule in (9).

(9) Ordering Rule:

Transition states are ordered: Continue is preferred to Retain, which is preferred to Smooth-shift, which is preferred to Rough-shift.

The Ordering Rule is intended to reflect the intuition that some transitions between utterances are more coherent than others. Maintaining the same topic (Cb) is more coherent than shifting topics, and maintaining the topic in subject position is more coherent than maintaining it in another position. Under the assumption that speakers try to maximize coherence, the prediction is made that continues should be the most frequent transition, that retains should be the next most frequent transition, then smooth shifts and lastly rough shifts. We will see later that this prediction does not entirely hold up. The ordering rule does work for continues, however; and it predicts certain cases of disambiguation, for example in (10):

(10) (i) entons x-Ø-ki-kanuj ru-beyal chi x-Ø-ki-lesaj ri-winaq,
then CPL-3sA-3pE-find 3sE-way that CPL-3sA-3pE-take.from def-people,

ri-por-taq-winaq ri-e-k'o chiri oher,
def-poor-plur-people def-3pA-exist there early

'Then they found a way to take it from the people, the poor people who had been there a long time.'

Cb: [the rich people]
Cf: [the rich people, the poor people, there]
Ct: [continue]

(ii) i x-e-ki-komsax

and CPL-3pA-3pE-kill

'And they killed them.'

Cb: [the rich people]

Cf: [the rich people, the poor people, there]

Ct: [continue]

Here, there are two possible interpretations for sentence (ii): either the rich people killed the poor people, or the poor people killed the rich people. Only the former is consistent with the ordering rule, however, since it would represent a continue transition, whereas the less preferred interpretation would represent a smooth shift.

2.7. Null Cb.

In addition to these explicit transitions, an utterance can have no Cb. Recall that, from the list of constraints in (2), Constraint 1 says there is at MOST one backward-looking center. Constraint 1 indicates that it is possible for an utterance to have no Cb. This generally happens when the utterance begins a new discourse segment, and shares no referents with the last utterance of the preceding segment.

Although, as we said earlier, Centering Theory is intended as a model of local coherence, in this paper, we will explore ways in which Centering Theory can provide clues to global discourse structure. One such way is utterances with no Cb. They indicate the opening of a new discourse segment. An example is shown in (11). The context here is that Enrique's sister is in the hospital and Enrique has a dilemma.

(11) no Cb

(i) porke ria k'o ni-Ø-ra-jo

because he exist INCPL-3sA-3sE-want

chi ni-Ø-pe-naj

that INCPL-3sA-come-far

chin-Ø-r-il ru-samaj

that INCPL-3sA-3sE-find 3sE-work

i n-Ø-u-chek juba ru-rajel.

and INCPL-3sA-3sE-earn some 3sE-money

'Because he wants to go far away to find work and earn some money.'

Cb: [Enrique]

Cf: [Enrique, work, money]

Ct: [continue]

(ii) i qatsi:j-nawe x-Ø-apon ri-ru-ræchbil ri-ya-ros

and truly CPL-3sA-arrive def-3sE-friend def-fem-Rosa

ri-ya-nache ru-bi.

def-fem-Nadia 3sE-name

'And, as it happened, the friend of Rosa's arrived, whose name was Nadia.

Cb: [?]

Cf: [Nadia, Rosa]

Ct: [?]

(iii) x-Ø-apon.

CPL-3sA-arrive

‘She arrived.’

Cb: [Nadia]

Cf: [Nadia]

Ct: [continue]

(iv) i x-Ø-u-bi:j che

and CPL-3sA-3sE-say to-3sE

chi rit ma-jun akuchi y-a-be-we

that you no-one where INCPL-2sA-go-place

‘and she said to him, “You are not going anywhere”.’

Cb: [Nadia]

Cf: [Nadia, Enrique]

Ct: [continue]

Here, the discourse had been about Enrique, the Cb in sentence (i), and it shifts to talk about a new referent, Rosa’s friend. Because Rosa’s friend had not been mentioned in the previous sentence, she can’t count as a backward-looking center in sentence (ii). The entrance of Rosa’s friend opens up a new discourse segment, which goes on to talk about her name and what she said to Enrique. Notice also that the new referent (Rosa’s friend) is introduced in post-verbal subject position, a position in Cakchiquel that is specialized for the introduction of new, important referents.

This ends the discussion of the theory. We turn now to a discussion of our analysis of our data.

3. The Data

Our method in analyzing the text was to code each sentence for its backward-looking center and list of forward-looking centers. We then calculated the transition types. The distribution of transition types is shown in Table 2:

continue	121	52.6 %
retain	41	17.8
smooth shift	49	21.3
rough shift	4	1.7
no CB	15	6.5
TOTAL	230	100 %

Table 2. Distribution of Transition Types

Note that the ordering rule is violated by these data: contrary to prediction, smooth-shifts occurred more often than retains. This pattern has also been found in other analyses of naturally occurring discourse (Hurewitz 1998: 279-280). Perhaps the ordering rule should simply be changed to prefer smooth shifts over retains. An explanation for this might be that subject position is the dominant position for backward-looking centers: Both Continues and Smooth-Shifts have the Cb in subject position whereas Retains and Rough Shifts have it in non-subject positions.

We now turn to our next step in our analysis of our data. Because we were interested in the distribution of different forms of referring expression, we coded each backward looking center for form of expression used to refer to it (zero, full pronoun and full NP). We then calculated the distribution of forms in each of the four transition types. This distribution is shown in Table 3.

	Ø	pro	np	Total	% Ø
continue	98	17	6	121	81%
retain	35	5	1	41	85
smooth shift	22	14	13	49	45
rough shift	3	0	1	4	75
Total	158	36	21	215	73%

Table 3. Distribution of Expression Forms by Transition Type

One interesting fact about this distribution is that backward-looking centers are usually encoded by zero anaphors. This happens 73% of the time.

It is also interesting that continues and retains are encoded with zeros more often than smooth shifts and rough shifts. Moving to Table 4, collapsing the full pronoun and full NP columns together and collapsing continues with retains and smooth shifts with rough shifts, we were able to perform a chi-square statistical test on the distribution. The results came out statistically significant to support the observation of a strong tendency to zero pronouns in continues and retains, whereas smooth and rough shifts were slightly more likely to use full pronouns or full NP's than zero pronouns.

	Ø	pro+np	Total
continue + retain	133	29	162
smooth+rough shift	25	28	53
Total	158	57	215

Table 4. Collapsed Distribution Table

$\chi^2 = 23.25$ (with Yates correction)

degrees of freedom = 1

$p < .001$

In sum, these patterns in Tables 2 and 3 indicate that center transitions can indeed partially explain our subject of principle interest: the distribution of different forms of referring expression in discourse: zeros are used for continuation of the previous Cb and fuller forms are used for shifts in Cb. A number of issues are still open, however, some of which relate to the other area of our interest—discourse segmentation. We now turn to a discussion of four of these.

4. Open issues

First, is the puzzle of why some continues and retains are realized as full pronouns or full NPs instead of zeros 29 out of 162 times in our sample (Table 4). Why would a speaker demote a referent to a form associated with less salience when it still continues to be the Cb? Our first hypothesis was that the fuller forms signal the beginning or end of a discourse segment. To test this hypothesis, we coded the sentences in the text as to whether they realize the beginning or end of an intuitive discourse segment. We then compared these codes with the form/transition correlations that we had already calculated in Table 3. What we found is that, for full NPs, five of the six continues from Table 3, or 83%, correlated with the beginning of a discourse segment. (12)(iii) shows an example of a full NP beginning a new discourse segment:

(12) (i) sino k'o jun mismo ri ki-winaq,
since exist one same def 3pE-people

jun señora ri kiere-desir
one woman def that-is-to-say

akwi ki-ch'alal ki-banon chi-la watemal.

how 3pE-relative 3pE-doing to-there Guatemala

'since there was one of their people, a woman, who that is to say, was just like their relative down there in Guatemala.'

Cb: [Rosa & Enrique]
Cf: [the woman, Rosa & Enrique, Guatemala]
Ct: [retain]

(ii) x-Ø-u-spaj juba poq chi-k-e
CPL-3sA-3sE-give some money to-3pE
'She gave some money to them.'

Cb: [the woman]
Cf: [the woman, Rosa & Enrique]
Ct: [smooth shift]

(iii) porke ri-jun-señora-re ru-yak'on juba ru-rajel
because def-one-woman-here 3sE-save some 3sE-money
'Because this woman had been saving her money'

Cb: [this woman]
Cf: [the woman, her money]
Ct: [continue]

(iv) ria x-Ø-ra-jo-ta x-Ø-pe-ta awe
def-3sA CPL-3sA-3sE-want-IRR CPL-3sA-come-IRR here

chuqa ri-estados-unidos oher

also def-states-united early

'She would have wanted to come here also to the U.S. before'

Cb: [the woman]

Cf: [the woman, the U.S.]

Ct: [continue]

(v) per majun-mod ma-x-Ø-pe-ta

but nothing-way neg-CPL-3sA-come-IRR

'But she didn't come.'

Cb: [the woman]

Cf: [the woman]

Ct: [continue]

The discourse-segment boundary here is between (ii) and (iii). The discourse segment shifts from talking about Rosa & Enrique's relationship to the woman—that she was their relative, that she gave them money—to talking about the woman herself and why she had saved up the money.

We decided that the one NP that we did not find correlated with a new discourse segment was articulated in order to disambiguate the utterance, as shown in (13):

(13) (i) i k'o chi jun ri-achin ri ma-qas-ta ni-Ø-r-qa-ta

and exist that one def-man who neg-almost-IRR INCPL-3sA-3sE-?- IRR

ch-Ø-wach

to-3sA-face

'And there was a man who didn't quite like him.'

Cb: [Enrique]

Cf: [a man, Enrique]

Ct: [retain]

(ii) achike ri-enrik komo uts

what def-Enrique as good

'Because Enrique was a good man.'

Cb: [Enrique]

Cf: [Enrique]

Ct: [continue]

Without the full NP in sentence (ii), the zero or full pronoun would be ambiguous as to whether it referred to the man who doesn't like Enrique or to Enrique.

Moving to full pronoun continues: interestingly, they seem to be partially correlated with discourse segment closings, although the results are not so clear here. We ignore first and second person pronouns here due to lack of time, and because direct discourse is not yet well understood in Centering Theory. 46% or 6 out of 13 full third-person pronouns indicate the end of a discourse segment. They seem to signal a kind of summing up of the discourse segment. An example is shown in (14):

(14) (i) i x-Ø-cha:p ru-ma r-ahaf ri-hayç ri-beyon-taq-winaq

and CPL-3sA-caught 3sE-by 3sE-owner def-house def-rich-plur-people
'And she was caught by the owner of the house of the rich people.'

Cb: [Rosa]
Cf: [Rosa, rich people, house]
Ct: [continue]

(ii) porke ria nabi?el xa x-Ø-u-ch'ej ri-tsi?eq ru-k'in ru-q'a
because sheat.first just CPL-3sA-3sE-wash def-clothes 3sE-with 3sE-hand
'because she at first just washed the clothes with her hands.'

Cb: [Rosa]
Cf: [Rosa, the clothes, her hands]
Ct: [continue]

(iii) i ki-ri x-Ø-u-ban aqal e-benaq-we
and thus CPL-3sA-3sE-do slow 3pA-doing-place
'And thus, they are doing well.'

Cb: [Rosa & Enrique]
Cf: [Rosa & Enrique]
Ct: [smooth shift]

Here, after a string of zeros referring to the Cb, Rosa, a full pronoun is used in sentence (ii). Sentence (iii) begins a new discourse segment, shifting the discussion from how Rosa was doing at her job, to how both brother and sister are doing in general. The full pronoun here seems to indicate a summing up of the discourse segment that it closes.

Another use of full pronouns is in providing a transition between a full noun phrase and a zero pronoun. An example can be seen in (15):

(5) (i) ri-ya-ros i ru-k'in ri-enrik ki-te ki-tata x-e-komses
def-fem-Rosa and 3sE-withdef-Enrique 3pE-mother 3pE-father CPL-3pA-killed
'Rosa and Enrique, their mother and father were killed'

Cb: [?]
Cf: [mother & father, Rosa & Enrique]
Ct: [?]

(ii) rie k'o ki-jyu, k'o ki-tiko?n chi-la watemal
they exist 3pE-land, exist 3pE-crops to-there Guatemala.
'They had land, they had crops down there in Guatemala.'

Cb: [mother & father]
Cf: [mother & father, land, crops, Guatemala]
Ct: [continue]

(iii) per ri-beyoma-taq-winaq ni-Ø-ki-jo ri-ki-jyu
but def-rich-plur-people INCPL-3sA-3pE-want def-3pE-land
'But the rich people wanted their land.'

Cb: [mother & father]
Cf: [the rich people, mother & father, land]
Ct: [retain]

This seems to occur in order to establish the new referent more securely before going on to discuss it.

So, in sum, we found that discourse segment boundaries and transitions account for 11 out of 13 continues with third-person pronouns, or 85%. This seems to be a good proportion of cases accounted for.

A second open issue is the function of retains in discourse. The original proposal in Centering Theory was that a retain corresponds to a situation where the speaker is intending to shift onto a new entity in the next utterance. However, only 20 out of 41 of the retains in our data, or 49%, were followed by a smooth or rough shift. The other retains were followed by a continue or another retain. A typical example of a retain followed by a continue is shown in (16)

(16) (i) x-e-ek'o-pe ch-u-pan ri-tonel
CPL-3pA-cross-come to-3sE-in def-tunnel
'They crossed inside the tunnel.'

Cb: [Rosa & Enrique]
Cf: [Rosa & Enrique, the tunnel]
Ct: [continue]

(ii) per ri-mas k'eyef ru-k'eyewal x-Ø-ki-wil pa-ki-beç
but def-most difficult 3sE-difficulties CPL-3sA-3pE-find in-3pE-way

ri-ch'oy-a x-e-k-il pa-beç
def-rat-plur CPL-3pA-3pE-find in-way

'But the most difficult of the difficulties they found on the way was the rats they found on the way.'

Cb: [Rosa & Enrique]

Cf: [the most difficult of the difficulties, the rats, Rosa & Enrique, the way]

Ct: [retain]

(iii) x-e-k-wil kongan ri-ch'oyç

CPL-3pA-3pE-find many def-rat

'They found many rats.'

Cb: [Rosa & Enrique]

Cf: [Rosa & Enrique, the rats]

Ct: [continue]

In sentence (ii), the speaker continues talking about the same participants--Rosa & Enrique--but shifts to a metalevel of the discourse—"the most difficult of the difficulties"—to frame the new discourse segment, which goes on to talk about how they crossed the Mexican-U.S. border.

An example of a retain followed by a retain is shown in (17). This one occurs at the very end of the narrative.

(7) (i) i x-Ø-chop ria samaj chi-la

and CPL-3sA-start he work to-there

'And he started working there.'

Cb: [Enrique]

Cf: [Enrique, there]

Ct: [continue]

(ii) i ru-nojel x-Ø-pe pa ru-jolon

and 3sE-all CPL-3sA-come in 3sE-head

akwi x-Ø-komses ru-tata

how CPL-3sA-killed 3sE-father

'And everything came to his mind how his father got killed'

Cb: [Enrique]

Cf: [everything, Enrique, mind, father]

Ct: [retain]

(iii) porke ru-tata no sol x-Ø-komses ru-ma eskopet,

because 3sE-father not only CPL-3sA-killed 3sE-by shotgun

ru-ma q'atbel-taq-tsij watemal

3sE-by authority-plural-word Guatemala

'Because his father not only got killed by a shotgun, by the Guatemalan army'

Cb: [Enrique]

Cf: [father, Enrique, shotgun, Guatemalan army]

Ct: [retain]

(iv) sino chuqa x-Ø-eleses ru-jolon

but also CPL-3sA-cut.off 3sE-head

'But also got his head cut off'

Cb: [father]

Cf: [father, head]

Ct:[smooth-shift]

(v) x-Ø-tse?qe ru-jolon chuqa jun che?

CPL-3sA-hang 3sE-head also one tree

'His head was hung on a tree.'

Cb: [father]

Cf: [father, head, tree]

Ct:[continue]

(vi) Matiox**.**

Thank.you

'Thank you'

Here the two retains in sentences (ii) and (iii) act together, this time in sequence. To shift the discourse segment, the speaker first makes a generalization—"everything"—and then becomes more specific—how his father got killed. This brings the father progressively into prominence where he can act as the backward-looking center.

In sum, as a result of a close examination of our data, it seems as if the generalization does hold that the function of retains tends to be to signal a shift, either to introduce a new referent or to introduce a new discourse segment.

The third open issue is how to determine the role of subordinate clauses of various sorts in determining center realization. We just adopted the simplest assumption that entities encoded in subordinate clauses are placed in subordinate position on the Cf list. We coded our narrative

sentence by sentence, or independent clause by independent clause. We didn't analyze dependent clauses separately. It does seem to us, however, that some dependent clauses do play a role in determining centers, especially initial adverbial clauses. However, we must leave the exploration of this issue to future research.

The fourth open issue is the question of how the form of referents which are not Cbs is determined. To some extent this issue is addressed by the Pronoun Rule in that non-Cbs can't have a less explicit form than the Cb. That is, if the Cb is realized as a full NP, other entities on the Cf list can't be realized as pronouns or zeros, and if the Cb is realized as a full pronoun, other entities on the Cf list can't be realized as zeros. But the theory doesn't predict cases where, for example, there are multiple zero pronouns in an utterance. Centering Theory only allows one discourse referent to be talked about at a time (either singular or plural), but it seems to us that sometimes more than one entity is in the focus of attention. A theory of the relative salience of discourse entities such as that of Gundel, Hedberg & Zacharski (1993) does allow for multiple in-focus entities (and hence multiple zero anaphors), but again we have to leave the issue of a comparison between the two theories to future research.

In our view, the best property of Centering Theory is that it is formalized precisely enough to enable many exact generalizations to be made about a large text such as the one we are working with. We look forward to continuing to explore how the theory can explain the structure of texts such as El Norte.

References

Brennan, Susan E., Marilyn Walker Friedman, and Carl J. Pollard

1987. A Centering Approach to Pronouns. *In* Proceedings of the 29th Annual Meeting of the Association of Computational Linguistics, 155-162. Stanford University.

Cote, Sharon.

1998. Ranking Forward-Looking Centers. *In* Centering Theory in Discourse, edited by Marilyn A.Walker, Aravind K. Joshi, and Ellen F. Prince, 55-69. Oxford: Clarendon Press.

Di Eugenio, Barbara.

1998. Centering in Italian. . *In* Centering Theory in Discourse, edited by Marilyn A.Walker, Aravind K. Joshi, and Ellen F. Prince, 115-157. Oxford: Clarendon Press.

Iida, Masayo.

1998. Discourse Coherence and Shifting Centers in Japanese Texts. *In* Centering Theory in Discourse, edited by Marilyn A.Walker, Aravind K. Joshi, and Ellen F. Prince, 161-180. Oxford: Clarendon Press.

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski.

1993. Cognitive Status and the Form of Referring Expressions in Discourse. Language, 69: 274-307.

Hurewitz, Felicia.

1998. A Quantitative Look at Discourse Coherence. *In* Centering Theory in Discourse, edited by Marilyn A.Walker, Aravind K. Joshi, and Ellen F. Prince, 273-291. Oxford: Clarendon Press.

Joshi, Aravind K. and Scott Weinstein.

1981. Control of Inference: Role of Some Aspects of Discourse Structure—Centering. Proceedings of the International Joint Conference on Artificial Intelligence, 385-387. Vancouver, B.C.

Rambow, Owen.

1993. Pragmatic Aspects of Scrambling and Topicalization in German: A Centering Approach. Paper presented at the Workshop on Centering Theory in Naturally-Occurring Discourse, Institute for Research in Cognitive Science, University of Pennsylvania, May 8-10.

Turan, Ümit Deniz.

1998. Ranking Forward-Looking Centers in Turkish: Universal and Language-Specific Properties. *In* Centering Theory in Discourse, edited by Marilyn A.Walker, Aravind K. Joshi, and Ellen F. Prince, 139-160. Oxford: Clarendon Press,

Walker, Marilyn A., Aravind K. Joshi, and Ellen F. Prince.

1998. Centering in Naturally Occurring Discourse: An Overview. *In* Centering Theory in Discourse, edited by Marilyn A.Walker, Aravind K. Joshi, and Ellen F. Prince, 1-28. Oxford: Clarendon Press,

Walker, Marilyn A. and Ellen F. Prince.

1996. A Bilateral Approach to Givenness: A Hearer-Status Algorithm and a Centering Algorithm. *In* Reference and Referent Accessibility, edited by Thorstein Fretheim and Jeanette K. Gundel, 291-306. Amsterdam: John Benjamins.

---

[1] Our consultant was Emilio Chagil, originally from San Lucas, Guatemala. He was the consultant in a field methods course at the University of Minnesota in 1981 and 1983-84. The first author attended both courses and afterwards continued working with him. The text was produced on 14 February, 1984. Emilio narrated, the first author transcribed and glossed the text, and Emilio translated it.

[2] The formulation of the theoretical rules and constraints given here is adapted from the presentation in Walker and Prince 1996, and from the introduction to Walker, Joshi and Prince 1998.

[3] The phonemic symbols should be interpreted as follows:

x	unvoiced palatal fricative
ch	unvoiced alveopalatal affricate
j	unvoiced velar fricative
q	unvoiced uvular stop
‘	glottalized consonant (ejective or implosive)

The abbreviated morphological glosses should be interpreted as follows:

CPL	completive aspect	IRR	irrealis
INCPL	incompletive aspect	def	definite

A	absolutive	MOD	modal
E	ergative	neg	negative
3	third-person	cont	continuative
s	singular	plur	plural
p	plural	fem	feminine

'Ø' indicates the third-person singular absolutive zero morpheme.

[4] Centering theorists have proposed that languages may differ in their ranking of forward-looking centers. For example, Iida 1998 proposes that topics and empathy in Japanese rank higher than subjects. Also, Rambow 1993 proposes that word order in German affects the Cf template (see Cote 1998 for discussion). In the absence of evidence to the contrary and for simplicity, we employ the ranking originally formulated by Brennan, Friedman and Pollard 1987.

[5] This formulation of the Pronoun Rule is based on Turan 1998, who proposes that zero subject pronouns in Turkish realize the Cb when they occur. Note that Cakchiquel permits zero subjects and zero objects as well as zero possessors and obliques.

[6] Here, we follow Di Eugenio 1998: 125 in ranking the possessor immediately following the possessee when the posessee is animate, and as immediately preceding the posessee when the possessee is inanimate.