

## Centering and noun phrase realization in Kaqchikel Mayan

Nancy Hedberg<sup>a</sup>

<sup>a</sup>Simon Fraser University, Linguistics Department, Burnaby, British Columbia, V5A 1S6, Canada

Email: hedberg@sfu.ca  
Tel: 1-778-782-3479  
fax: 1-778-782-5659

### **Abstract:**

Centering Theory is applied to a narrated film retelling in Kaqchikel Mayan in order to better understand discourse constraints on the form of referring expression. It is shown that Backwards Looking Centers are very often encoded by zero pronouns, and that center Shifts more often employ full pronouns and full noun phrases than do center Continues and Retains. Preverbal pronouns and full noun phrases as opposed to postverbal pronouns and full noun phrases correlate with center Shifts. Morphological ergativity is seen to have no effect on the proper ranking of Forward Looking Centers in that in morphologically ergative languages, just as in morphologically accusative languages, subjects should be taken to outrank objects in transitive sentences. A comparison of the Kaqchikel text with DuBois' (1987) analysis of Sacapultec Mayan film-retelling narratives supports this ranking by showing that ergative subject arguments more often encode participants that are 'given' in the discourse than do absolutive object arguments.

**Keywords:** Reference, Centering Theory, Pronoun, Zero Pronoun, Ergativity, Mayan languages

**References from abstract:** Du Bois, John W., 1987. The discourse basis of ergativity. *Language* 63 (4), 805-855.

### **Bio-note:**

Nancy Hedberg is an Associate Professor in the Department of Linguistics at Simon Fraser University in Canada. Her research interests include syntax, semantics, pragmatics and cognitive science. Her current work focuses on cleft sentences, form of referring expression, and the meaning of prosody.

## 1. Introduction

Centering Theory is a theory of local discourse structure. It can be seen as an algorithm for tracking discourse entities throughout a text, or as an algorithm for determining text coherence. Unlike many approaches to discourse analysis that arise from anthropological linguistics and sociolinguistics, Centering Theory is a theory of discourse coherence originating in computational linguistics. It was developed in Grosz, Joshi & Weinstein 1983, 1995; Brennan, Friedman & Pollard 1987; Walker, Joshi & Prince 1998; as well as much further work.

Centering Theory has been applied to English, German, Dutch, Yiddish, Italian, Spanish, French, Greek, Polish, Hindi, Turkish, Japanese, Korean, Chinese, and Yapese (see Wiesemann (forthcoming) for references). However, never before has Centering Theory been applied to an aboriginal language of the Americas, nor to a morphologically ergative language, except insofar as Hindi shows ergativity. The goal of the present paper is to fill these gaps by presenting a Centering Theory analysis of an oral narrative in Kaqchikel Mayan.

First, I'll present an introduction to the Kaqchikel language, the narrative text that I have analyzed, and Kaqchikel grammatical structure. Then I'll introduce Centering Theory. I will then discuss the results of the Centering theory analysis of the Kaqchikel text, focusing on insights that Centering Theory can give us with regard to the structure of Kaqchikel discourse, especially with regard to predicting the form of referring expression used. Finally, I'll conclude by comparing the Kaqchikel narrative with the analysis of Du Bois 1987 of oral narratives in the closely-related Sacapultec Mayan language, with a view to uncovering factors that shed further light on the proper role of ergativity in Centering Theory analyses.

### 1.1. Kaqchikel

Kaqchikel is principally spoken in Guatemala. It has about half a million speakers, and is a member of K'ichee' branch of Mayan languages. The orthography was devised by the Academy of Mayan Languages of Guatemala.<sup>1</sup>

---

<sup>1</sup> I have made use of a recent textbook/grammar/dictionary, McKenna Brown et al. (2006).

The phonemic symbols should be interpreted as follows:

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

The abbreviated morphological glosses should be interpreted as follows:

CPL	completive aspect	IRR	irrealis
INCPL	incompletive aspect	DEF	definite
A	absolutive	TR	transitive
E	ergative	NEG	negative
3	third-person	CONT	continuative
s	singular	PLUR	plural
p	plural	FEM	feminine
CAUS	causative	PASS	passive

‘Ø’ indicates the third-person singular absolutive zero prefix.

The narrative retells the story of the feature film, *El Norte*, which was made in 1984 with Gregory Nava as director. The film tells the story of a Mayan brother and sister who leave Guatemala after their parents are killed, immigrate illegally to the U.S. and try to survive there. The narrative lasts 14 minutes and comprises 365 clauses.<sup>2</sup>

Kaqchikel is a radically head-marking language (Nichols 1986). Verbs are marked for the person and number of their subject and object. Prepositions (or relational nouns) are marked for the person and number of their object. And nouns are marked for the person and number of their possessor. The language is massively pro-drop: Cross-referenced verbal subjects and objects as well as possessors and many prepositional objects are omissible when sufficiently salient in the discourse, but can also appear as overt pronouns or full noun phrases.

Kaqchikel is morphologically ergative in its agreement system. Subjects of transitive verbs are crossreferenced with ergative agreement markers on the verb. Objects of transitive verbs and subjects of intransitive verbs are crossreferenced with absolutive agreement markers on the verb, possessors of noun are marked with ergative markers on the noun, and objects of prepositions are marked ergative on the preposition. The word order is often SVO, but sometimes VSO, VOS or OVS, and even SOV or OSV if the verb is preceded first by a topic and then by a focus—with the focus antipassive required in case of OSV. Importantly, subjects can appear in either preverbal or postverbal position. There is no gender distinction in the agreement prefix system, and there are no gender or case distinctions in the pronoun system.

There are a number of marked syntactic constructions, the most frequently occurring one being the passive, where the patient is promoted to subject and marked absolutive, and the agent is demoted to oblique. There is also the antipassive, where the agent is marked absolutive and the patient is demoted to oblique; and the focus antipassive, whereby in agent questions, relative clause, and clefts, only the agent is marked on the verb with the absolutive marker, and the patient is demoted to oblique.<sup>3</sup> Kaqchikel also exhibits right and left dislocation.

## 1.2 Centering Theory

I now turn to a brief introduction to Centering Theory as I have applied it. First, I will present an introduction, then an illustration of the transitions, a description of the two rules of Centering Theory, and finally a description of how I segmented sentences into utterances.

Each utterance introduces a set of discourse entities. Each discourse entity is a member of a ranked set of Forward-Looking Centers, or Cf list. One member of the Cf list is the center of attention, termed the Cb, or the Backwards-Looking Center. Coherent texts will relate discourse entities across utterances in an orderly fashion. Thus, the Cb of an utterance is the highest ranked Cf of the previous utterance that is realized in the current utterance. The Preferred Center or Cp is the highest ranked entity in the current utterance. It is a constraint of Centering Theory that there is at most one Cb per utterance, but an initial utterance or an utterance that doesn't share any arguments with the previous utterance has no Cb.

---

<sup>2</sup>The narrative was collected during a 1984 Field Methods class taught by Nancy Stenson at the University of Minnesota. The class and the consultant, Emilio Chagil, went to see the film, *The next day*, the consultant retold the story of the film. The text was transcribed and glossed by Nancy Hedberg and Emilio Chagil in 1984, and was translated by Emilio Chagil. It was analyzed initially for Centering Theory in 1999 by Nancy Hedberg and Sandra Dueck, with a revision made in 2007 by Nancy Hedberg, with help from Maite Taboada.

<sup>3</sup>This aspect of the syntax of Kaqchikel and other Mayan languages has been analyzed as an instance of syntactic ergativity (Du Bois 1987, *inter alia*). The choice of argument to receive the absolutive agreement on the verb in the focus antipassive is actually more complex, and involves a person hierarchy.

The Forward-Looking Centers of an utterance are taken to be salient to different degrees, and this is assumed to be a result of how they are expressed grammatically. The grammatical salience system is understood to vary cross-linguistically, so that different languages have different Cf Templates (Cote 1998), but typically, subjects are taken to outrank objects, which outrank other entities. For possessive noun phrases, I followed Di Eugenio 1998 in analyzing possessor as outranking possessed if the possessed is inanimate; and possessed as outranking possessor if the possessed is animate.

There are four types of Center Transitions (Ct) between utterances, as shown in Table 1, depending upon whether the two utterances share the same Cb or not, and whether the Cb of the second utterance is also the Cp. If the two utterances share the same Cb or if the first utterance has no Cb, then we have a Continue transition if the Cb of the second utterance is the Cp, and we have a Retain if the Cb is not the Cp. If the Cb is different in the second utterance, then we have a Smooth Shift if the Cb of the second utterance is the Cp, and we have a Rough Shift otherwise.

	$Cb(U_i) = Cb(U_{i-1})$	$Cb(U_i) \neq Cb(U_{i-1})$
	$Cb(U_{i-1}) = \text{NONE}$	
$Cb(U_i) = Cp(U_i)$	<b>CONTINUE</b>	<b>SMOOTH-SHIFT</b>
$Cb(U_i) \neq Cp(U_i)$	<b>RETAIN</b>	<b>ROUGH-SHIFT</b>

Table 1. Center Transitions (Walker, Joshi and Prince 1998)

Example (1) from the El Norte text shows a Continue in (ii), followed by a Retain in (iii), followed in turn by a Smooth Shift in (iv).

- (1) (i) k'eyef-b'a x-ø-ki-b'än  
 difficult-some CPL-3sA-3pE-do  
 'They got into problems.'  
 Cb: Rosa & Enrique  
 Cf: Rosa & Enrique difficulty  
 Ct: CONTINUE
- (ii) i yamer x-e-käm-s-ës pa-b'ey ru-ma jun k'ajol-achin  
 and almost CPL-3pA- die-CAUSE-PASS on-road 3sE-by one young-man  
 aj-mejikan  
 PERSON-Mexican  
 'And they were almost killed on the road by a young man, a Mexican'  
 Cb: Rosa & Enrique  
 Cf: Rosa & Enrique, road, Mexican  
 Ct: CONTINUE
- (iii) porke x-ø-u-b'ij chi-k-e chi ria y-e-r-q'asaj-pe  
 because CPL-3sA-3sE-say to-3pE-to that s/he INCPL-3pA-3sE-take-come

kere-pe estads-unids

here-come states-united

‘Because he said to them that he would take them across over here to the United States.’

Cb: Rosa & Enrique

Cf: Mexican, Rosa & Enrique, U.S.

Ct: RETAIN

- (iv) pero ma-n-Ø-u-ya-ta ri-ru-pensar  
 but NEG-INCPL-3sA-3sE-give-IRR DEF-3sE-thought  
 ‘But that wasn’t his intention [He didn’t give his thought.]’  
 Cb: Mexican  
 Cf: Mexican, his thought  
 Ct: SMOOTH SHIFT

In (ii) Cb = Cp, and since the Cb didn't change, the transition is a Continue. The Cb again stays the same in (iii) but Cb ≠ Cp, so the transition is a Retain. Finally in (iv) the Cb shifts to the previous Cp and Cb = Cp, so we have a Smooth Shift. It has been suggested that a sequence of Continue, Retain and Smooth Shift is a maximally coherent way to shift Cb (Grosz et al 1995, Kibble 2001).

For Rough Shift, we have to turn to a different part of the story, as shown in (2). The Cb changes from Rosa to Enrique, but the Cb of the second utterance is not the Cp, so we have a Rough Shift. Note that there were only three rough shifts in the entire discourse.

- (2) (i) kwan enrik x-Ø-u-tzat r-ana’ chi ria xa najin kam  
 when Enrique CPL-3sA-3sE-see 3sE-sister that s/he just CONT die  
 ‘When Enrique saw his sister that she was dying’  
 Cb: Rosa  
 Cf: Enrique, Rosa  
 Ct: RETAIN
- (ii) i x-Ø-kam r-ana’  
 and CPL-3sA-die 3sE-sister  
 ‘And his sister died.’  
 Cb: Enrique  
 Cf: Rosa, Enrique  
 Ct: ROUGH SHIFT

There are two rules of Centering theory. The first is the Pronoun Rule, which for English states that if any member of the set of Forward Looking Centers is expressed by a pronoun, the Cb must be. This rule captures the insight that the Cb will be expressed with the shortest, most coherent referring expression of that utterance. Because Kaqchikel allows zero and full pronouns in both subject and object position as well as zero and full pronoun possessors, Hedberg &

Dueck 1999 adapted the formulation of the pronoun rule in (3) from Turan 1998, who had generalized the Pronoun Rule to account for null subjects in Turkish.<sup>4</sup>

(3) **Pronoun Rule (Kaqchikel):**

In each  $U_i$  in a discourse segment  $D$  consisting of utterances  $U_1, \dots, U_n$ , 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  $Cb(U_i)$ .

The second rule is the Ordering Rule, which asserts that transition states are ordered, as shown in (4).

(4) **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. The ordering rule predicts certain cases of disambiguation, for example in (5):

- (5) (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 who 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 (Ø<sup>5</sup>)  
 Cf: the rich people, way, land, the poor people, there  
 Ct: continue

- (ii) i x-e-ki-kam-s-ax  
 and CPL-3pA-3pE-die-CAUSE-TR  
 ‘And they killed them.’

Cb: the rich people (Ø)  
 Cf: the rich people, the poor people  
 Ct: continue

<sup>4</sup> This formulation of the Pronoun Rule is consistent with the extended Pronoun Rule proposed by Strube and Filippova 2007.

<sup>5</sup> The 'Ø' here indicates that this argument was not lexically expressed as either a full or pronominal noun phrase.

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.

Two final issues have to do with segmentation. First, complex sentences must be separated into utterances. Utterance segmentation has been discussed by Kameyama 1998 and Miltsakaki 2003, among others. Here, I treated adverbial clauses as separate utterances, and I treated relative clauses and complement clauses as part of the main clause utterance. Continuations of direct discourse complement clauses were treated as separate utterances. Second, Centering Theory is a theory of local coherence and presupposes a theory of global coherence that involves organizing a text into discourse segments. While I make some reference to discourse segments, a treatment of discourse segmentation is beyond the scope of this paper. The coding was done as if the entire text were one segment.

## 2. Results and Discussion

### 2.1 Centering Theory Analysis of the El Norte Narrative

I now turn to the results of the Centering Theory analysis of the narrative. I'll discuss three primary results. The first involves the correlation between transition type and the form of the referring expression denoting the Cb. The second involves preverbal and postverbal realizations of Cbs. The third has to do with morphological ergativity.

#### 2.1.1. *Transition and Form of Referring Expression.*

Table 2 shows the form in which the Cb was realized under transitions of the different types. It can be seen that the raw numbers of transitions of the different types follows the Ordering Rule. There are more Continues than Retains, more Retains than Smooth shifts<sup>6</sup>, and more Smooth Shifts than Rough Shifts. Also it can be seen that the Cb is likely to be expressed with a zero pronoun, especially when the transition is a Continue.<sup>7</sup> When full NPs are used, the transition is most likely a Smooth Shift; and full pronouns are an intermediate case, being used either for Continues or Smooth Shifts.<sup>8</sup>

---

<sup>6</sup> A number of researchers have found that Smooth Shifts outnumber Retains in texts, especially when EST is considered a separate transition (c.f. footnote 8). See Kibble 2001 and Taboada & Wieseemann (this volume) for references and discussion. Kibble presents a reformulation of the Ordering Rule of Centering Theory to accommodate this fact. Note that here, Retains have a greater proportion of zero pronoun Cb realizations (84.1%) as Continues (82%), while Smooth Shifts have far fewer (35.1%). This might constitute evidence that the original version of the Ordering Rule is the correct one.

<sup>7</sup> Taboada & Wieseemann (this volume) found that the preferred realization of Cbs in English telephone conversations is a pronoun and the preferred realization of Cbs in Spanish telephone conversations is a zero pronoun.

<sup>8</sup> 'ZERO' is the label from Poesio et al. 2004 for a transition between an utterance that has a Cb to one without a Cb. These authors use the label 'NULL' for a transition from one utterance without a Cb to another without a Cb, and the label 'EST' (Establishment) for a transition between an utterance without a Cb to one with a Cb. No NULL transitions were found in my data, and I treated EST transitions as Continues or Retains, as per Walker, Joshi and Prince 1998.

	NONE	Ø	Pronoun	NP	Demon Pronoun	TOTAL
CONT		123	19	6	2	150
RET		37	5	2		44
SS		13	8	16		37
RS		3				3
ZERO	23					23
TOTAL	23	176	32	24	2	257

Table 2: Cb Realizations under Transitions of Different Types

Table 3 compares zero pronouns with personal pronouns and lexical noun phrases, across three transition types: Continue, Retain and Smooth Shift.

	Ø	PRONOUN PERS or DEMON	FULL NP
CONT	71.1% (123)	61.8% (21)	25.0% (6)
RET	21.4% (37)	14.7% (5)	8.3% (2)
SS	7.5% (13)	23.5% (8)	66.7% (16)
	100% (173)	100% (34)	100% (24)

Table 3: Table of Cb Realizations and Transitions

Notice that zero pronouns are used primarily for continues (71.1%), and lexical NPs are used primarily for smooth shifts (66.7%). Full pronouns are intermediate, being used for continues 61.8% of the time and for smooth shifts 23.5% of the time. This behavior of full pronouns merits examination since full pronouns give no more information than the verbal prefixes that cross-classify them, so why would full pronouns be used?

An examination of particular examples shows that full pronouns are often used to unambiguously convey smooth shift over continue. An example is given in (6).

- (6) (i) i k'o chi jun señor ri naj ni-Ø-pe-we.  
 and exist that one man that far INCPL-3sA-come-WE<sup>9</sup>  
 'And there was a man who came from far away.'  
 Cb: NONE  
 Cf: man, far away  
 Ct: ZERO

<sup>9</sup> -we is a suffix that occurs on a verb when a locative has been fronted or extracted from a sentence.

- (ii) x-Ø-u-suj ru-samaj ri-enrik.  
 CPL-3sA-3sE-offer 3 sE-work DEF-enrique  
 'He offered Enrique work.  
 Cb: the man (Ø)  
 Cf: the man, Enrique, work  
 Ct: CONTINUE
- (iii) pero **ria** k'o juba ru-pena n-Ø-u-ya-qa ri-ya-ros  
 but he exist some 3sE-pain INCPL-3sA-3sE-leave-down DEF-FEM-Rosa  
 'But he is afraid to leave Rosa behind.'  
 Cb: Enrique (ria)  
 Cf: Enrique, pain, Rosa  
 Ct: SMOOTH SHIFT

Here, the full pronoun in the third utterance shifts the Cb from the man to Enrique. If a zero had been used, it might have sounded like the man was going to continue to be the Cb. Thus, use of an overt pronoun helps to signal the shift. Four out of eight full pronoun Smooth Shifts were used to disambiguate the utterance like this. The other four cases involved lack of potential ambiguity due to person differences (the pronoun is first or second person).

This result suggests that full pronouns should be placed in the Activated position of a Givenness Hierarchy for Kaqchikel (Gundel, Hedberg & Zacharski 1993), while zero pronouns would be placed in the In Focus position. Full pronouns in Kaqchikel function like neuter demonstrative pronouns in English in signaling a focus shift, as in (7), when an unstressed *it* would signal focus continuation.<sup>10</sup>

- (7) When Snepp makes a speech he has to submit a text to CIA censors first. When he wrote a book review for the Los Angeles Times, he had to show it to the agency before he sent it to the newspaper, and when the editor asked for a change, he had to show **that** to the censors too. (Gundel, Hedberg, & Zacharski 1993, example (56)).

The disambiguating full pronoun Smooth Shift cases should be compared with the 13 zero pronoun cases. Is ambiguity tolerated in the case of zero pronouns? It turns out that 11 out of these 13 cases are disambiguated syntactically: two by differential number marking, six by differential person marking, two by being continuations of direct discourse, and one by taking part in a parallel construction. It thus appears that expressing a referent by means of a full pronoun is indeed a way of disambiguating a discourse reference when a Cb Smooth Shift is taking place.

An example showing how ambiguity is alleviated by use first of an overt pronoun and then by a full NP is given in (8). The speaker here first used the full pronoun strategy, but then shifted to a fully disambiguating full noun phrase.

- (8) (i) k'o ki-tiko'n chi-la watemal  
 exist 3pE-crops to-there Guatemala  
 'They had crops down there in Guatemala.'

<sup>10</sup> Gundel, Hedberg and Zacharski 1993 also posit that unstressed personal pronouns need to be in focus while stressed personal pronouns only need to be activated.

Cb: Mother & Fahter (Ø)  
 Cf: Mother & Father, crops, Guatemala  
 Ct: CONTINUE

- (ii) per ri-beyoma-taq-winäq 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: Mother & Father, rich people, land  
 Ct: RETAIN

- (iii) porke **rie** x-Ø-ki-ts'at, **ri-beyoma-taq-winäq** x-Ø-ki-ts'at  
 because they CPL-3sA-3pE-see, DEF rich-PLUR-people CPL-3sA-3sE-see  
 'because they saw, the rich people saw'

chi ri-jyu uts-laj-jyu uts-laj-tiko'n.  
 that DEF-land good-LAJ-land good-LAJ-crops.  
 'that the land was good land, good crops.'

Cb: rich people (NP)  
 Cf: rich people, land  
 Ct: SMOOTH SHIFT

Here the Cb is Enrique's and Rosa's mother and father (i). This Cb is retained as Cb in (8ii) and the rich people are introduced as Cp. The Cb shifts to the rich people in (iii), with the speaker first using a pronoun to effect this shift but then shifting to a full noun phrase to fully disambiguate.

### 2.1.2 Preverbal versus Postverbal Realization

Notice that the full pronouns in (6) and (8) occurred in preverbal position and signaled a Smooth Shift. Table 4 shows that there is a correlation between preverbal versus postverbal realization of the Cb and center transition. Both preverbal and postverbal position can be used to realize center Continue or Retain (55.9% vs. 44.1%), but Smooth Shifts are likely to be realized preverbally (70.8%).

	Pre-verbal	Post-verbal	
CONT+RET	55.9% (19)	44.1% (15)	100% (34)
SS	70.8% (17)	29.2% (7)	100% (24)

Table 4: Preverbal vs. Postverbal Cb Realizations and Transitions

Example (8) is another example showing a preverbal pronoun signaling a Smooth Shift, from the immigration authorities as Cb in (9i) back to Rosa and Enrique as Cb in (9ii).

- (9) (i) x-e-ki-jo-ta                      x-e-ki-tzolij-el  
 CPL-3pA-3pE-want-IRR CPL-3pA-3pE-send-leave  
 ‘They wanted to send them back.’  
 Cb: Authorities (Ø)  
 Cf: Authorities, Rosa & Enrique  
 Ct: CONTINUE
- (ii) per komo **rie** ki-bän chel kom mejikan  
 but as they 3pE-do like as Mexican  
 ‘But they made like Mexicans.’  
 Cb: Rosa & Enrique (pre-verbal rie)  
 Cf: Rosa & Enrique, Mexicans  
 Ct: SMOOTH SHIFT

Again if a zero pronoun had been used, it might have been taken as re-referring to the immigration authorities.

Example (10) shows a postverbal pronoun signaling a center Continue.

- (10) (i) x-e-käm-s-ës-ta  
 CPL-3pA-die-CAUS-PASS-IRR  
 ‘[because if they had stayed in Guatemala,] they would be killed.’  
 Cb: Rosa & Enrique (Ø)  
 Cf: Rosa & Enrique  
 Ct: CONTINUE
- (ii) aqal x-Ø-ki-tamaj                      **rie** akwi ni-Ø-ki-bän                      chi y-e-ek'ase                      awe  
 slowly CPL-3sA-3pE-learn they how INCPL-3pA-3pE-do to INCPL-3pA-survive here  
 ‘Slowly they learned how to survive here’  
 Cb: Rosa & Enrique (post-verbal rie)  
 Cf: Rosa & Enrique, here  
 Ct: CONTINUE

Here (10i) presents an utterance with Rosa and Enrique as Cb. (10ii) continues to talk about Rosa and Enrique, and the Continue Cb is realized as a postverbal pronoun. Perhaps an overt pronoun instead of a zero is selected here because the utterance begins a new discourse segment.<sup>11</sup>

### 2.1.3 Morphological Ergativity

In ordering the set of forward-looking centers, I have made use of grammatical relations, with subject outranking object. This means that a morphologically ergative language like Kaqchikel is given the same treatment as a morphologically accusative language like English. That is, the

<sup>11</sup> (10i) ends a long discourse segment about about how Rosa and Enrique got to the United States. The pitch trails off at the end of (10i) and the speaker's voice gets very soft in intensity. There is a long pause, and then a reset to a noticeably higher pitch at the beginning of (10ii). For evidence that such prosodic characteristics signal discourse segment boundaries, see Hirschberg and Grosz (1992).

subject of an intransitive verb will be the Cp, and in a transitive sentence, the subject will outrank the object. Might it not be the case, however, that morphological case should be taken as a factor in ranking the set of forward-looking centers, so that morphologically ergative languages have a different ranking system than morphologically accusative languages?

To test this hypothesis, some portions of the text were reanalyzed in such a way that absolutive object arguments were taken to outrank ergative subject arguments. The question is whether Subject over Object or Absolutive over Ergative leads to more coherent transitions. In testing this hypothesis, it readily becomes evident that the SUBJ > OBJ method of calculating Cf order is superior to an ABS > ERG method. That is, there are many complement clauses in the data, which are cross-referenced on the main clause verb as third person singular absolutive arguments. An example is shown in (11).

- (11) (i) entons ri-enrik i ya-ros x-Ø-ki-ts'at  
 then DEF-Enrique and FEM-Rosa CPL-3sA-3pE-see
- ri-ki-te' ki-tata x-e-käm-s-ës.  
 DEF-3pE-mother 3pE-father CPL-3pA die-CAUS-PASS  
 'Then Enrique and Rosa saw their mother and father killed.'
- Cb None  
 Cf Enrique & Rosa, Mother & Father  
 Ct ZERO
- (ii) x-Ø-ki-xbij-ki  
 CPL-3sA-3pE-scare-3pE  
 'They were scared.'
- Cb: Enrique & Rosa (Ø)  
 Cf: Enrique & Rosa  
 Ct: CONTINUE

Here, if ABS > ERG order were followed, Mother and Father would rank higher than Enrique and Rosa on the Cf list of (i) since the subject (only argument) of the absolutive subordinate clause would outrank the ergative argument of the main clause. Since Mother and Father are animate possessed arguments, they would outrank the possessor (and ergative) Enrique and Rosa. However, Enrique and Rosa are the zero pronoun referents of the only argument of the following sentence (ii), so the transition would have to be a Retain on the ABS > ERG analysis (since the Cp would be Mother and Father); whereas the more coherent Continue transition can be assigned under the SUBJ > OBJ analysis. Since it is frequently the case that the ergative subject of the main clause is referred to again in immediately subsequent discourse, as in this example, it is clear that this argument should not be ranked lower than all of the entities introduced in the subordinate clause.

Table 5 shows a side-by-side comparison of one discourse segment that does not involve complement clauses. The more coherent transition is in bold face and underlined.

**ERG SUBJ > ABS OBJ**

**ABS OBJ > ERG SUBJ**

And also they needed to borrow money

<u>CONTINUE</u>	RETAIN
Thanks to God, they didn't have to borrow the money.	
<u>CONTINUE</u>	SMOOTH SHIFT
Since there was one of their people, a woman, who that is to say was just like their relative down there in Guatemala.	
RETAIN	RETAIN
She gave some money to them.	
SMOOTH SHIFT	<u>RETAIN</u>
Because this woman had been saving money.	
<u>CONTINUE</u>	SMOOTH SHIFT
She would have wanted to come here also to the United States before.	
<u>CONTINUE</u>	SMOOTH SHIFT
But in any case she didn't come.	
CONTINUE	CONTINUE
Then she gave it to them, Rosa and Enrique.	
<u>CONTINUE</u>	RETAIN

Table 5: Center transitions under two alternative orderings of ERG and ABS arguments.

It turns out in this discourse segment that  $SUBJ > OBJ$  ranking was more coherent five times, the  $ABS > ERG$  ranking was more coherent only once, and the two rankings gave the same result twice. It can be concluded that the ranking based on grammatical relations works best for Kaqchikel, and that morphological ergativity, therefore, does not affect the Cf Template. That is, subjects should outrank objects in an ergative-absolutive system just as they do in a nominative-accusative system.

## 2.2 Ergativity and Discourse Structure

It may be the case that a Cf ranking hierarchy based on grammatical relations works best in morphologically ergative as well as morphologically accusative languages because an ergative marked argument (the subject) is more salient than an absolutive argument (the object) in transitive sentences in discourse. Motivation for this conclusion comes from two sources. First, from a Centering Theory perspective, Prasad & Strube (2000) found that transitive subjects in Hindi outrank objects with respect to the Pronoun Rule, despite Hindi showing ergativity in its pronoun system. Secondly, Du Bois (1987), from an anthropological linguistics perspective, shows that in Sacapultec Mayan narrative discourse, ergative arguments are restricted to be expressed using zero arguments or pronominal arguments and are almost exclusively 'given' in terms of information status.<sup>12</sup> Full lexical noun phrases overwhelmingly tend to only appear in absolutive position, and it is these positions that host 'new' arguments. In this section I compare Du Bois' findings with regard to his Sacapultec narratives with the El Norte narrative in Kaqchikel. Like me, Du Bois investigated film-retelling narratives, specifically retellings of the Pear Story film of Chafe (1980).

<sup>12</sup> Sacapultec (Sakapultek), also spoken in Guatemala, is closely related to Kaqchikel, both being members of the K'ichee' branch of Mayan languages.

Du Bois (1987) found that new information is presented in intransitive subject (S) and transitive object (O) position—i.e. positions corresponding to absolutive case; while subjects of transitive verbs (A) overwhelmingly present given information—i.e. positions corresponding to ergative case. "Given" here means Discourse Old (Prince 1992). Du Bois terms this information structure pattern the "Given A Constraint". We also see this constraint at work in the Kaqchikel narrative, as Table 6 shows for both languages (Du Bois' Table 6).<sup>13, 14</sup>

	GIVEN		NEW		Total	
	N	%	N	%	N	%
<u>Sacapultec</u>						
A	180	96.3	7	3.7	187	100
S	187	72.5	71	27.5	258	100
O	111	65.3	59	34.7	170	100
Total	478	77.7	137	22.3	615	100
<u>Kaqchikel</u>						
A	153	96.8	5	3.2	158	100
S	107	88.4	14	11.6	121	100
O	67	66.4	34	33.7	101	100
Total	327	86.1	53	13.9	380	100

Table 6. Grammatical relation and information status in Sacapultec and Kaqchikel.

A arguments in Kaqchikel as well as in Sacapultec overwhelmingly tend to be given (for Sacapultec,  $\chi^2 = 56.4$ , d.f. = 2,  $p < .001$ ; for Kaqchikel,  $\chi^2 = 48.6$ , d.f. = 2,  $p < .001$ ). S arguments tend to be given more often in the Kaqchikel than in the Sacapultec data, probably as a reflection of the fact that the Kaqchikel data comes from a single narrative that continues to talk about the same characters, while the Sacapultec data comes from a set of 18 shorter narratives, each of which introduces the same set of characters, often using S arguments to do so.<sup>15</sup>

This information structure difference between A, O and S arguments is correlated with differences in the forms of referring expression used in the three positions in extended discourse. Du Bois classified the three types of arguments found in his Pear Story narratives into Lexical (i.e. full noun phrases), Pronominal, and Zero ("Affixal") morphological types, and extracted the distribution shown in top half of Table 7 (his Table 2). The Kaqchikel distribution is shown in the bottom half of the table.<sup>16</sup>

<sup>13</sup> I have classified Du Bois' few 'accessible' referents together with 'new' as this is how I coded the Kaqchikel data.

<sup>14</sup> It isn't clear why the Sacapultec numbers in Table 7 are different from the numbers in Table 6. The numbers are different in DuBois' Tables 2 and 6.

<sup>15</sup> The Kaqchikel narrative also supports Du Bois' 1987 "One New Argument Constraint". Out of 101 clauses containing both an A and an O argument, only one clause had both a new A argument and a new O argument.

<sup>16</sup> Like Du Bois (1987), I didn't count equative clauses and relative clauses. I also didn't count indirect questions.

	ZERO		PRONOMINAL		LEXICAL		Total N
	N	%	N	%	N	%	
<u>Sacapultec</u>							
A	156	86.7	13	7.2	11	6.1	180
S	124	47.3	12	4.6	126	48.1	262
O	94	53.1	2	1.1	81	45.8	177
Total	374		27		218		619
<u>Kaqchikel</u>							
A	103	65.2	29	18.4	26	16.5	158
S	73	60.3	15	12.4	33	27.3	121
O	27	26.7	3	3.0	71	70.3	101
Total	203		47		130		380

Table 7. Grammatical relation and morphological type of mention in Sacapultec and Kaqchikel.

It can be seen in Table 7 that in Sacapultec, A arguments tend to be expressed as zero pronouns, while S and O arguments can be expressed either lexically or as zeros. Full pronouns are quite rare in the Sacapultec narratives. The Kaqchikel narrative follows a similar pattern in that A arguments tend to be expressed as zero pronouns, while O arguments tend even more than in Sacapultec to be lexically expressed. S arguments in Kaqchikel tend to be expressed more often with zero pronouns than in the Sacapultec data, again probably because the same small set of characters continues to be talked about throughout the long narrative. Also, full pronouns are used more often than in the Sacapultec data, perhaps for the same reason.

In sum, this section has shown that A arguments, i.e. subjects of transitive verbs, which are marked with ergative agreement markers on the verb in Mayan languages, are indeed more often given than O arguments, i.e. objects of transitive verbs. This suggests that even in morphologically ergative languages, where S and O have the same marking, subjects of transitive verbs should be ranked higher than objects on the Cf Hierarchy in Centering Theory, supporting the conclusion reached in §2.1.3 on the basis of the relative coherence of centering transitions. This information status pattern predictably correlates with form of referring expression in that A arguments are less likely to be lexically expressed than O arguments.

### 3. Conclusion

We have seen that Centering Theory can be used to track main characters through a narrative. In Kaqchikel, zero pronouns are used when the Cb of the current utterance and the Cb of the previous utterance are the same, and full pronouns and NPs are used for Cb shifts. In addition, preverbal pronouns and NPs are used for Cb shifts and postverbal ones are used for Cb

continuation. Ergative arguments in transitive sentences can be argued on a number of grounds to outrank transitive absolutive arguments in terms of discourse salience, so the grammatical relations Cf ranking method used for nominative-accusative languages (subject > object) can also be applied to morphologically ergative languages. Centering Theory is useful in that it provides a precise algorithm for tracking the most salient referent in a text. It can predict when a form fuller than a zero pronoun should be used to refer to a discourse entity at the center of attention. In this way, it can explain why full pronouns are used more often in subject position than object position, even though subject position is a more salient position.

It is still the case, however, that more general theories of referring expression realization like the Givenness Hierarchy framework of Gundel, Hedberg and Zacharski (1993) and the Accessibility Theory of Ariel (1990) are needed to account for the selection of form of referring expression in positions that are not relevant to the issue of competing for Cb status. Thus, for example, full pronouns as opposed to full lexical noun phrases are sometimes used in salient positions (e.g. subject) in utterances that have no Cb because they share no entities with the previous utterance. The Givenness Hierarchy theory and Accessibility Theory would predict that such pronouns should be acceptable when the referent is a salient discourse topic, but Centering Theory is silent about these cases, which involve reference to discourse topics that extend across discourse segments. Centering theory is only intended as a theory that tracks referents within discourse segments. In addition, referring expressions still need to be chosen in non-Cb positions of sometimes quite long sentences. As long as the Pronoun rule is obeyed, Centering Theory does not constrain such choices. Despite these limitations, the present study has shown that Centering Theory can explain many aspects of the use of different forms of referring expression in the most salient positions in an attested narrative of a novel language.

**ACKNOWLEDGEMENTS.** I would like to thank Maite Taboada for help in the reanalysis of the El Norte text with respect to Centering Theory. A version of this paper was presented at the 10<sup>th</sup> International Pragmatics Association Conference in the panel on Reference organized by Thorstein Fretheim, Kaja Borthen and Heidi Brøseth. I am grateful to the organizers for a stimulating panel, and to members of the audience for useful feedback. I also thank two anonymous reviewers for helpful criticism. An early version of this paper was published as Hedberg & Dueck 1999. I thank Ellen Prince and Donna Gerdts for helpful comments on that paper, as well as audiences at the 1999 WSCLA conference and the 2000 LSA conference. The research for the earlier paper was supported by SSHRC Grant #410-94-1081, and the present revision was supported by SSHRC grant #410-2007-0345. Finally, I gratefully acknowledge Emilio Chagil for producing, translating and discussing the narrative that I have analyzed, and Nancy Stenson for facilitating the original research.

## References

- Ariel, Mira, 1990. *Accessing Noun Phrase Antecedents*. Routledge, London.
- 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*. Stanford University, pp. 155-162.
- Chafe, Wallace L. (Ed.), 1980. *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Ablex, Norwood, NJ.

- Cote, Sharon, 1998. Ranking forward-looking centers. In: Walker, M.A., Joshi, A. K., Prince, E. F. (Eds.), *Centering Theory in Discourse*, Clarendon Press, Oxford, pp. 55-69.
- Du Bois, John W., 1987. The discourse basis of ergativity. *Language* 63 (4), 805-855.
- Di Eugenio, Barbara, 1998. Centering in Italian. In: Walker, M.A., Joshi, A. K., Prince, E. F. (Eds.), *Centering Theory in Discourse*, Clarendon Press, Oxford, pp. 115-157.
- Grosz, Barbara J., Joshi, Aravind K., Weinstein, Scott, 1983. Providing a unified account of definite noun phrases in discourse, *Proceedings of the 21<sup>st</sup> Annual Meeting of the Association for Computational Linguistics*.
- Grosz, Barbara J., Joshi, Aravind K., Weinstein, Scott, 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21 (2), 203-225.
- Gundel, Jeanette K., Hedberg, Nancy, Zacharski, Ron, 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69 (2), 274-307.
- Hedberg, Nancy, Dueck, Sandra, 1999. Cakchiquel reference and Centering Theory. *Proceedings of the Workshop on Structure and Constituency in the Languages of the Americas*, University of British Columbia Working Papers in Linguistics, 59-74.
- Hirschberg, Julia and Barbara Grosz, 1992. Intonational features of local and global discourse structure, *Proceedings of the DARPA Workshop on Spoken Language Systems*, Arden House.
- Kameyama, Megumi, 1998. Intrasentential centering: A case study. In: Walker, M.A., Joshi, A. K., Prince, E. F. (Eds.), *Centering Theory in Discourse*, Clarendon Press, Oxford, pp., 89-112.
- Kibble, Roger, 2001. A reformulation of Rule 2 of Centering Theory. *Computational Linguistics* 27 (4), 579-587.
- McKenna Brown, R., Maxwell, Judith M., Little, Walter E., 2006. *¿La üt awäch? Introduction to Kaqchikel Maya Language*. University of Texas Press, Austin.
- Miltsakaki, Eleni, 2003. *The Syntax-Discourse Interface: Effects of the Main-Subordinate Distinction on Attention Structure*. PhD dissertation, University of Pennsylvania.
- Nichols, Johanna, 1986. Head-marking and dependent-marking grammar. *Language* 62.56-119.
- Poesio, Massimo, Stevens, Rosemary, Di Eugenio, Barbara, Hitzeman, Janet, 2004. Centering: A parametric theory and its instantiations. *Computational Linguistics* 30 (3). 309-363.
- Prasad, Rashmi, Strube, Michael, 2000. Discourse salience and pronoun resolution in Hindi. *University of Pennsylvania Working Papers in Linguistics*, 6 (3), 189-208.
- Prince, Ellen F. 1992. The ZPG letter: Subjects, definiteness, and information status. In: Mann, W.C., Thompson, S.A. (Eds.), *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, John Benjamins, Amsterdam, pp. 295-325.
- Strube, Michael, Filippova, Katja, 2007. Constraints on pronominalization: Extending the coverage of Centering's Rule 1. Paper presented at the 10<sup>th</sup> International Pragmatics Association Conference, Göteborg, Sweden, July 8-13, 2007.
- Taboada, Maite, Wiesemann, Loreley, this volume. Subjects and topics in conversation. *Journal of Pragmatics*.
- Turan, Ümit Deniz, 1998. Ranking forward-looking centers in Turkish: Universal and language-specific properties In: Walker, M.A., Joshi, A. K., Prince, E. F. (Eds.), *Centering Theory in Discourse*, Clarendon Press, Oxford, pp. 139-160.
- Walker, Marilyn A., Joshi, Aravind K., Prince, Ellen F., 1998. Centering in Naturally Occurring Discourse: An Overview In: Walker, M.A., Joshi, A. K., Prince, E. F. (Eds.), *Centering Theory in Discourse*, Clarendon Press, Oxford, pp., 1-28.

Wiesemann, Loreley Marie, forthcoming. The Function of Spanish and English Relative Clauses in Discourse and their Segmentation in Centering Theory. PhD dissertation, Simon Fraser University.