

## **Anaphoric terms and focus of attention in English and Spanish\***

Maite Taboada

The choice of one anaphoric term (pronoun, clitic, noun phrase) over another is related to what is in the focus of attention at each point in the discourse (the entities and relations mentioned by the participants in a conversation). In this paper, I explore the relation between choice of anaphoric expression and focus of attention by applying a theory of local focus in discourse, Centering Theory (Grosz, Joshi, & Weinstein 1995). I examined spoken language corpora in English and Spanish, and determined the relationship between the focus of attention and the type of anaphoric term used for the topic of each utterance. Results show that focus is not the only factor involved; in conversation, other factors, such as

**UNCORRECTED PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

as zero anaphora, whereas least accessible topics are represented as indefinite noun phrases. In general, the encoding and presentation (or ‘packaging’) of different types of discourse referents is the object of a number of studies (Chafe 1976; Daneš 1974; Kuno 1972; Lambrecht 1994; Prince 1981; Vallduví 1990; Vallduví & Engdahl 1996).

Centering Theory builds a hierarchy of entities in the discourse, for each discourse segment. The entities are ranked according to different criteria, grammatical function being the preferred (for English at least). For each discourse segment, one entity is

the intentions, or purposes, of the discourse participants, (ii) the attention of the participants and (iii) the structure of the discourse. Centering is concerned with the participants' attention and how the global and local structures of the discourse af-

UNCORRECTED PROOFS  
© JOHN BENJAMINS PUBLISHING COMPANY

Table 1. Transition types

	$Cb(U_i) = Cb(U_{i-1})$ or $Cb(U_{i-1}) = \emptyset$	$Cb(U_i) \neq Cb(U_{i-1})$
$Cb(U_i) = Cp(U_i)$	$\langle \cdot, \cdot, \cdot, \cdot \rangle$	$\langle \cdot, \cdot, \cdot, \cdot \rangle$
$Cb(U_i) \neq Cp(U_i)$	$\langle \cdot, \cdot, \cdot, \cdot \rangle$	$\langle \cdot, \cdot, \cdot, \cdot \rangle$

be concerned with that topic, the  $Cp(U)$  in a  $\langle \cdot, \cdot, \cdot, \cdot \rangle$ ; or it may be not linked to the previous topic, in a  $\langle \cdot, \cdot, \cdot, \cdot \rangle$ . Transitions are one explanation<sup>4</sup> for how coherence is achieved: a text that maintains the same centers is perceived as more coherent.

In example (1), the first utterance has no Cb, because it is segment-initial, and therefore it has no transition (also called a zero-Cb transition). The transition between (1a) and (1b) is a  $\langle \cdot, \cdot, \cdot, \cdot \rangle$ , because the Cb of (1a) is empty, and the Cp and Cb of (1b) are the same,  $\langle \cdot, \cdot, \cdot, \cdot \rangle$ . Finally, the transition between (1b) and (1c) is a  $\langle \cdot, \cdot, \cdot, \cdot \rangle$ , since the Cb has changed. This signals a shift, the change to a new Cb ( $\langle \cdot, \cdot, \cdot, \cdot \rangle$ ), possibly the referent that is going to be continued, since it is also the Cp. Because transitions capture topic shifts in the conversation, they are ranked according to the demands they pose on the reader. The ranking is:  $\langle \cdot, \cdot, \cdot, \cdot \rangle > \langle \cdot, \cdot, \cdot, \cdot \rangle > \langle \cdot, \cdot, \cdot, \cdot \rangle$ . This transition ranking is often referred to as Rule 2 in the Centering paradigm. Centering predicts that  $\langle \cdot, \cdot, \cdot, \cdot \rangle$  will be preferred to  $\langle \cdot, \cdot, \cdot, \cdot \rangle$ , and  $\langle \cdot, \cdot, \cdot, \cdot \rangle$  to  $\langle \cdot, \cdot, \cdot, \cdot \rangle$ , all other things being equal. The preference applies both to single transitions and to sequences of transitions.

Rule 1 captures the preference for pronouns when the same topic of discourse is continued. The formulation of Rule 1 is as follows:

For each  $U$  in a discourse segment  $D$  consisting of utterances  $U_1, \dots, U_n$ , if some element of  $Cf(U, D)$  is realized as a pronoun in  $U$ , then so is  $Cb(U, D)$ .

Rule 1 is sometimes referred to as the *pronoun preference rule*. It captures the fact that a topic that is continued from a previous utterance does not need to be signalled by more explicit means than a pronoun (or a zero pronoun, in languages that allow those). Other pronouns are of course allowed in the same utterance, but the most salient entity must

### 3. Applying Centering to spoken language

Centering has been, for the most part, applied to written language. A number of issues arise as we extend it to cover spoken language. These have to do with the interpretation of discourse segment or utterance; with spoken language phenomena such as false starts, repetition, overlapping and backchannel signals; and with the treatment of first and second person pronouns, which are not usually considered as part of the study of discourse anaphora, but which are prevalent in spoken conversation.

Segmentation of spoken discourse is often fraught with difficulties. Speech does not happen in clear, period-separated sentences, but in smaller intonation units, some



**UNCORRECTED PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

The criteria for ranking the Cf list have long been debated. Originally, in a formulation that considered only English, the proposed ranking was based on grammatical relations: Subjects are more salient than Objects (Grosz et al. 1995). As other languages were studied, language-specific characteristics forced a revision of the criteria. For Japanese, topic markers and empathy became part of the ranking (Walker et al.

UNCORRECTED PROOFS  
© JOHN BENJAMINS PUBLISHING COMPANY



‘it seems to me’). With those verbs, the thematic role of Experiencer is ranked higher than the grammatical function of Subject.

- (8) me vien-e mejor el jueves,  
 I:1 come-3 better the Thursday  
 ‘Thursday is better for me.’ Cf: I ( ), (the meeting, null),

Animacy also affects the ranking: animate entities are ranked higher. Animacy interacts with empathy, since the Experiencer in psychological verbs always carries animacy. Animacy also interacts with linear order, since when both Indirect and Direct Object are expressed through clitics, in Spanish the Indirect Object (typically animate) is placed first. This hierarchy is already present in Givón’s (1983) studies on topic, which include semantic roles (and their grammatical function correlates) and animacy in calculations of topicality. As an example of animacy, we can see in (9b) that the indirect clitic ‘to her’ precedes the direct ‘it’, which refers to a scholarship for a program that was given to the speaker’s sister. The null Subject is arbitrary (it has no definite referent), and thus ranked last.

- (9) a. Mi hermana solicitó un programa de arqueología y antropología en Grecia.  
 ‘My sister applied to a program in archaeology and anthropology in Greece.’  
 b. ¡Y que se lo dan!  
 and that I:3 M:3 give:3  
 ‘And they gave it to her!’  
 Cf: ( , ‘to her’), ( , ‘it’), (null)

A more detailed description of all the phenomena pertaining to ranking the Cf list can be found in Hadic Zabala and Taboada (2004) and in Taboada (forthcoming). The order for Spanish is summarized below (10).

- (10) Experiencer > Subj > Animate IObj > DObj > Other > Impersonal/Arbitrary pronouns

##### 5. Choice of anaphoric term for the most salient entity

The purpose of this paper is to investigate what referring expression is chosen most often for the backward-looking center (Cb) in any given utterance, comparing English and Spanish. The hypothesis is that the choice of referring expression will depend on the type of transition holding between current and previous utterance. Some of the research in Centering has established certain trends. For instance, Di Eugenio (1998) showed that, typically, when the transition between  $U_1$  and  $U_2$  is a




**Table 3.** Transition types per language

	Continue	Retain	Smooth shift	Rough shift
English (n = 779)	65.6%	17%	14.4%	3%
Spanish (n = 790)	65%	16%	15%	4%

**Table 4.** Referring expressions for the Cb of each utterance; percentages are with respect to the transition type in each language

	Continue	Retain	Smooth shift	Rough shift
--	----------	--------	--------------	-------------

(13)  .

B: a. yo, yo, yo laburo ahora con la máquina del tipo éste, viste  
'I, I, I work now with this guy's computer, you know.'

A: b. sí, ¿qué tiene?

An interesting aspect in English is the presence of some zero pronouns, as in ex-

UNCORRECTED PROOFS  
© JOHN BENJAMINS PUBLISHING COMPANY

- B: c. No, no, ni ella lo podía creer,  
'No, no, not even she herself could believe it.'

Another interesting question that affects both languages is the choice of full noun phrases. Both use an NP to encode the Cb, even in cases where the Cb is the same as in the previous utterance (of all the  $\langle \text{Cb} \text{ NP} \rangle$  transitions, 12% in English; 16% in Spanish). We included proper names together with NPs, and many of the repetitions have to do with the use of proper names. Proper name repetition might be a device to establish common ground between the interlocutors. Downing (1996) points out that proper names are used very often in conversation: to introduce individuals in the conversation, as the most easily identifiable form of reference; and to refer again to those individuals, as a marker of true familiarity with the referent denoted by the proper name. In example (21),  $\langle \text{Cb} \text{ NP} \rangle$  (probably speaker A's partner, friend or child) has been part of the discourse, but never before mentioned by name. Speaker A talks about camping and biking, but she always uses the first person plural pronoun  $\langle \text{Cb} \text{ P} \rangle$  ('we camped there', 'we rented a bike'). Speaker B mentions the name  $\langle \text{Cb} \text{ NP} \rangle$ , part of the previous 'we', for the first time here, and speaker A repeats the name. It is difficult to explain why the proper name is repeated. B already knows that A is familiar with the referent, since  $\langle \text{Cb} \text{ P} \rangle$  was in A's party. It is possible that A repeats it to convey that she understood who B was referring to.

- (21) B: a. so did Johnny stay in front?  
A: b. Johnny was basically in front

### 5.3 $\langle \text{Cb} \text{ P} \rangle$ transitions

The percentage of pronoun forms (whether zero, pronoun, or clitic) in the  $\langle \text{Cb} \text{ P} \rangle$  transitions decreases considerably, as compared to the  $\langle \text{Cb} \text{ NP} \rangle$  transitions. In a  $\langle \text{Cb} \text{ P} \rangle$  there is a slight change of topic: the Cb of the current utterance is the same as the Cb of the previous utterance. However, the current utterance has seemingly moved the focus from the Cb to another entity, which is now the highest-ranked entity in the Cf list, typically the grammatical Subject. Example (22) shows the shift from  $\langle \text{Cb} \text{ NP} \rangle$  to I, the speaker. Since  $\langle \text{Cb} \text{ P} \rangle$  is the link between (22a) and (22c), it is the Cb (the backchannel in (22b) is ignored for analysis purposes). But in (22c) a new entity, the Subject, is the Cp, the highest ranked member of the Cf list.

- (22) A: a. the public beach is kind of hard to find  
Cf:  $\langle \text{Cb} \text{ NP} \rangle$  - Cb:  $\langle \text{Cb} \text{ P} \rangle$   
B: b. oh really  
A: c. I mean I didn't I didn't know where it was  
Cf:  $\langle \text{Cb} \text{ P} \rangle$  - Cb:  $\langle \text{Cb} \text{ P} \rangle$

The Cb may, in  $\langle \text{Cb} \text{ P} \rangle$  transitions, still be expressed through a pronoun. According to Di Eugenio (1998), this is possible especially in cases where the pronominalized

**UNCORRECTED PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**UNCORRECTED PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY



**UNCORRECTED PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

are also investigating optimal utterance segmentation methods, taking into account Centering structures, but also agreement among coders. Another area to explore is the relationship between the local focus of discourse, which Centering is supposed to handle, and global discourse structure (turn-taking, side sequences, overall discourse segments).

### Notes

\* A previous version of this paper was presented at the Third International Contrastive Linguistics Conference held in Santiago de Compostela (Spain) in September 2003. I would like to thank Loreley Hadic Zabala for helping in the coding of the data. This work was supported by the Ministry of Science and Technology of Spain, and the Xunta de Galicia, under project MCYT-FEDER BFF2002-02441/XUGA-PGIDIT03PXIC20403PN, and by Simon Fraser University, under a SSHRC grant and a Discovery Parks grant.

---

**References**

- Ariel, Mira (1990). *Anaphora in discourse*. London: Routledge.
- Bentivoglio, Paola (1983). "Topic continuity and discontinuity in discourse: A study of spoken Latin-American Spanish". In T. Givón (Ed.), *Reference and generality: A discussion of some medieval and modern theories* (pp. 255–311). Amsterdam and Philadelphia: John Benjamins.
- Brennan, Susan E. (1995). "Centering attention in discourse". *Journal of Pragmatics*, 17, 137–167.
- Brennan, Susan E., Marilyn W. Friedman, & Carl J. Pollard (1987). "A Centering approach to pronouns". In *Proceedings of the 2nd Annual Meeting of the Berkeley Linguistics Society* (AL-8) (pp. 155–162). Stanford, CA, USA.
- Brown, Gillian & George Yule (1983). *Discourse analysis*. Cambridge: Cambridge University Press.

UNCORRECTED PROOFS



Walker, Marilyn A., Aravind K. Joshi, & Ellen F. Prince (1998). "Centering in naturally occurring

UNCORRECTED PROOFS  
© JOHN BENJAMINS PUBLISHING COMPANY