# 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 turn-taking, seem to have a role in the type of anaphoric term chosen.

## 1. Aims and background

A basic question in discourse analysis, and in the study of anaphoric terms is: what is the anaphoric term chosen at any particular point in the discourse? The question encompasses a number of other issues that relate the choice of a particular anaphoric term (referring expression, discourse anaphora) to the context of situation and the preceding text. A speaker chooses a pronoun, a proper name, a definite noun phrase, an indefinite noun phrase, or a number of other possible types of noun phrase, depending on different factors: whether the referent for the anaphoric term has been mentioned before in the discourse; whether it was mentioned as a Subject or an Object; whether there are competing referents in the discourse; or what is most salient at that particular point in the discourse. The list could be much longer. The difficulty lies precisely in determining what are the factors that affect the choice of anaphoric term.

In this paper, I apply a theory of focus of attention in discourse, Centering Theory (Grosz et al. 1995), to the problem of determining what anaphoric term will be chosen for a particular referent. A number of other explanations have been proposed, based on a hierarchy of givenness of referents (Gundel et al. 1993), similar to a scale of accessibility of the referents encoded by different referring expressions (Ariel 1990), or based on the syntactic function of the referring expression (Gordon et al. 1999). Givón (1983) also correlates the form of the referring expression with topicality (what is in the focus of attention) in his topic hierarchy. For instance, most accessible topics are encoded

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 considered to be the topic of the discourse segment. This paper explores the choice of anaphoric term for that particular topic, which is always an entity that has been mentioned earlier in the discourse.

This study considers choices in two languages, English and Spanish. English has been widely studied in general theories of anaphora, and within Centering in particular (Brennan 1995; Brennan et al. 1987; Cornish 1999; Gordon et al. 1993; Hudson-D'Zmura & Tanenhaus 1998; Hurewitz 1998). A number of other languages have been studied using Centering as a framework: Italian (Di Eugenio 1998), Japanese (Iida 1998; Kameyama 1985; Walker et al. 1994), Turkish (Turan 1995), Greek (Miltsakaki 2001), German (Strube & Hahn 1996), Hindi (Prasad & Strube 2000). Spanish, however, has not received enough attention. The comparison between English and Spanish is interesting, since Spanish is different in a number of parameters (presence of *pro-drop* and clitics, more flexible word order). I examined spoken language corpora, which have often been neglected in discourse-based studies. Spoken language will provide more interesting insights for phenomena relating to the focus of attention, since attention issues are much more pressing in spoken language.

The corpus chosen for this study is the *CallHome corpus*, for English and Spanish. CallHome was an effort by the Linguistics Data Consortium to collect spontaneous telephone conversations. Participants were given thirty minutes of long-distance calling time, to call relatives or friends, provided they agreed to being recorded. For this particular study, I chose five conversations in English and five in Spanish.

The paper is organized as follows: in Section 2, I provide a brief overview of Centering Theory, the framework of the study. Since I am applying Centering to two relatively new areas, I discuss the process of applying Centering to spoken language in Section 3, and extending Centering to analyze Spanish in Section 4. Section 5 presents the main results of the study: the relationship between transition type and anaphoric term used for the *topic* of the sentence.[1] The findings are presented in terms of tendencies: what anaphoric term tends to be used, given a particular transition. Speakers often choose a term that does not follow those general tendencies. I discuss some of those cases also in Section 5, which is followed by conclusions in Section 6.


## 2.   Centering Theory

Centering (Grosz et al. 1995; Walker et al. 1998) was developed within a theory of discourse structure (Grosz & Sidner 1986) that considers the interaction between (i)

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 affect the referring expressions and the overall coherence of the discourse. It models the structure of local foci in discourse, i.e., foci within a discourse segment.

Centers are semantic entities that are part of the discourse model of each utterance in the segment. For each utterance, Centering establishes a ranked list of entities mentioned or evoked, the *forward-looking center list* (Cf). The list is ranked according to salience, defined most often in terms of grammatical relations (see Section 3). The first member in the Cf list is the *preferred center* (Cp). Additionally, one of the members of the Cf list is a *backward-looking center* (Cb), the highest-ranked entity from the previous utterance that is realized in the current utterance.

Example (1) below illustrates these concepts. Let us assume that the three utterances in the example constitute a discourse segment. In the first utterance of the segment, (1a), there are three centers: *Bilbo, Frodo* and *a ring*. In (1b), the centers are still the same, now all realized as pronouns (*he=Bilbo, him=Frodo, it=ring*). The list of forward-looking centers for (1a) includes all three centers, ranked according to grammatical function: Subject > Indirect Object > Direct Object.[2] The Cf list for (1a) is: Bilbo, Frodo, ring.[3] The preferred center in that utterance is the highest-ranked member of the Cf list, i.e., Bilbo. Similarly, in (1b), the Cf list contains, in this order: Bilbo, Frodo, ring. The Cp is still Bilbo. Notice that (1a) does not have a backward-looking center (the center is empty), because this is the first utterance in the segment. The backward-looking center of (1b) is Bilbo, the highest-ranked entity from (1a) that is also realized in (1b). In (2), we see the Cf, Cp and Cb for each of the three utterances in the segment.

(1)  a.  Bilbo gave Frodo a ring.
     b.  He told him to look after it very well.
     c.  Frodo put the ring in his pocket.

(2)  a.  Cf: Bilbo, Frodo, ring – Cp: Bilbo – Cb: Ø
     b.  Cf: Bilbo, Frodo, ring – Cp: Bilbo – Cb: Bilbo
     c.  Cf: Frodo, ring, pocket – Cp: Frodo – Cb: Frodo

In addition to the three types of centers, Centering proposes different types of transitions, based on the relationship between the backward-looking centers of any given pair of utterances, and the relationship of the Cb and Cp of each utterance in the pair. Transitions, shown in Table 1, capture the introduction and continuation of new topics. $Cb(U_i)$ and $Cp(U_i)$ refer to the centers in the current utterance. $Cb(U_{i-1})$ refers to the backward-looking center of the previous utterance. Thus, a continue occurs when the Cb and Cp of the current utterance are the same and the Cb of the current utterance is the same as the Cb of the previous utterance. Transitions capture the different types of ways in which discourse can progress: to give just two examples at opposite ends of the spectrum, an utterance may refer to a previous topic, the $Cb(U_{i-1})$, and still

**Table 1.** Transition types

|  | $Cb(U_i) = Cb(U_{i-1})$ **or** $Cb(U_{i-1}) = \varnothing$ | $Cb(U_i) \neq Cb(U_{i-1})$ |
|---|---|---|
| $Cb(U_i) = Cp(U_i)$ | CONTINUE | SMOOTH SHIFT |
| $Cb(U_i) \neq Cp(U_i)$ | RETAIN | ROUGH SHIFT |

be concerned with that topic, the $Cp(U_i)$ in a CONTINUE; or it may be not linked to the previous topic, in a ROUGH SHIFT. Transitions are one explanation[4] 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 CONTINUE, because the Cb of (1a) is empty, and the Cp and Cb of (1b) are the same, BILBO. Finally, the transition between (1b) and (1c) is a SMOOTH SHIFT, since the Cb has changed. This signals a shift, the change to a new Cb (FRODO), 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: CONTINUE > RETAIN > SMOOTH SHIFT > ROUGH SHIFT. This transition ranking is often referred to as Rule 2 in the Centering paradigm. Centering predicts that CONTINUE will be preferred to RETAIN, and RETAIN to SHIFTS, 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 $U_i$ in a discourse segment D consisting of utterances $U_1, \ldots, U_m$, if some element of $Cf(U_{i-1}, D)$ is realized as a pronoun in $U_i$, then so is $Cb(U_i, D)$.

Rule 1 is sometimes referred to as the *Pronoun 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 be realized by the least marked referring expression. In (1b), the backward-looking center, BILBO, is realized as a pronoun, following Rule 1, since a pronoun is also used to refer to FRODO, which is not the backward-looking center.

There is a relationship between the type of transition that occurs between any two utterances, and the type of anaphoric term chosen to realize entities in the focus of attention. Centering predicts that entities higher up in the focus (the Cf list) will be realized through zero pronouns, full pronouns or clitics, depending on the options available in the language. CONTINUE transitions, because they keep the same center, will likely encode the Cb or the Subject of the sentence as a zero pronoun or a full pronoun (Di Eugenio 1998). Shifts (smooth or rough) will result in less pronominalization. We will see that these relationships are quite complex, and different factors come into play in the choice of anaphoric term.

### 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 of which are not grammatically complete. Schiffrin (1994) argues for abandoning the sentence as a unit of analysis, and considering 'utterance' as the basic unit of organization in spoken language. However, the definition of utterance has not been fully spelled out. Even within Centering Theory, regardless of whether the language is written or spoken, some argue for a main clause with all its embedded clauses (Miltsakaki 2002);[5] some argue for the finite clause as the unit of analysis (Kameyama 1998); and some consider that most clauses, finite or not, should be units, or *utterances*, in the Centering sense (Poesio 2000).

We are in the process of evaluating different segmentation methods, and have already proposed a method (Hadic Zabala & Taboada 2004). Our current approach follows Kameyama (1998). It considers the tensed clause as the unit of analysis (the utterance, in Centering terms): conjoined clauses are separated as two (or more) units; main clauses are separated from any embedded clauses, provided the embedded clauses are finite; non-finite clauses form a unit with their matrix clause. In (3), two utterances are present, since (3b) is a tensed embedded clause. There is a third clause, *irme a ver a mi hermana* ('to go see my sister'), which is not an utterance, since the verb in non-finite.

(3)  a.   No compro nada, no nada, nada
'(I) don't buy anything'
b.   porque quiero irme a ver a mi hermana
'because (I) want to go see my sister.'

The next issue in spoken language is the presence of false starts, repetitions, backchannels and overlapping. Since Centering processes sentences in a linear order, these are important aspects to consider. False starts are utterances that are abandoned in favour of a different wording, or that may not be completed at all. Since some false starts can be units in themselves, we decided to consider as units only those that had entities in them, i.e., nominal expressions with reference (Eckert & Strube 1999). Centering is concerned with how entities are managed in the focus of discourse. As a consequence, we need to consider only those units that contain entities. Repetitions are treated similarly: only repeated units that contain entities are considered. Other repetitions are not included in the Centering analysis.

Backchannels (or backchannel signals) are expressions by the hearer that indicate to the speaker that the channel is still open, and that the communication is being successful (Yngve 1970). Backchannels are especially important in telephone conversations, which are the types of conversations analyzed. As with false starts, backchannels are discarded when they do not contain any entities. In example (4), speaker B provides a backchannel (*oh yeah*) to what speaker A is saying. The only units considered for the analysis are (4a) and (4c).

(4)  A:  a.  and we got one tree left over by the garbage can
     B:  b.  oh yeah
     A:  c.  yeah {laugh} so you got to leave one you know

Overlapping is particularly difficult, given the linear constraints in Centering. As a general rule, we have maintained the linearity in time reflected in the CallHome corpus time stamps. Each utterance in the corpus transcripts has a time stamp at the beginning, indicating how long into the conversation it was produced. We used those to order utterances, and also listened to the recordings. In some cases, it is clear to the researcher that speakers do not produce speech in a perfect linear fashion. Moreover, it is clear that even speech produced linearly has intricate connections. In example (5), Speaker A describes when he goes to buy dinner. He says, however, 'lunch', for food that he buys at 7 p.m. Speaker B corrects him, and asks another question about friends. Speaker A accepts the correction, and answers the question at the same time. If we look at the time stamps (the time interval when the utterance started and finished), we can see that Speaker B started to say *la cena* ('dinner') at 201.01. Speaker A started to accept that correction at 201.72. But while Speaker A was saying *Sí, la cena, claro*, Speaker B was already asking the next question. From a discourse anaphora point of view, we would want to be able to relate the entities in utterances (5a), (5b), and (5d) on the one hand; and (5c) and (5e), on the other. From a strictly linear point of view, however, utterances (5c) to (5e) have no entities connecting them.

(5)  193.92–200.51  A:  a.  No, a las siete, a las siete, siete y pico, suelo ir con un japonés y con un francés a comprar la comida allá enfrente.
                            'No, at seven, at seven, at seven something, (I) usually go with a Japanese (man) and a French (man) to buy lunch there across.'

     201.01–204.43  B:  b.  La cena.
                            'Dinner.'

                        c.  ¿Qué tal los amiguetes?
                            'How about friends?'

     201.72–211.02  A:  d.  Sí, la cena, claro.
                            'Yes, dinner, of course.'

                        e.  Bien, hay bastante gente de España,
                            'Good, there are quite a few people from Spain,'

Centering, like most theories of anaphora, is concerned with third person reference. First and second pronouns are deictic, referring to the speaker and hearer in the present situation, and not to a series of entities brought into the conversation by the speakers. Theoretically, then, first and second pronouns should not be part of a Centering analysis. Byron and Stent (1998) devised an experiment, also using part of the English CallHome corpus, to test the parameters of Centering. They found that a configuration of Centering that included first and second person pronouns performed better than an analysis using exclusively third person pronouns. The measures in an optimal configuration of Centering include: the number of empty-backward looking centers (Cb) in utterances;[6] whether the Cb coincided with the sentence topic; and whether the method in general produced more cheap or more expensive transitions. Cheap and expensive refer to the inference load placed on the hearer (Strube & Hahn 1999), based on whether the preferred center of an utterance ($Cp_U$), which is the expected Cb of the next utterance, is actually realized as $Cb_{U+1}$. We also decided to include first and second person pronouns in our analysis.

In summary, the approach taken here to spoken discourse uses strict linearity to relate utterances, disregards utterances that do not contain any entities (false starts, backchannels), and includes first and second person pronouns in the Cf list, although strictly speaking they are not anaphoric, but exophoric or deictic.

## 4.   Ranking entities in the focus of attention

Centering Theory, which derives from a theory of focus of attention and discourse structure (Grosz & Sidner 1986), attempts to describe the management of entities and topics in the discourse. It is concerned with which entities are part of the focus of the discourse, and which of those entities is more salient. Using that knowledge, one could predict facts about reference and pronominalization. For these purposes, Centering has created a number of constructs.

Traditionally, the most salient entity in the discourse has been called the discourse topic, with the sentence topic being roughly the counterpart at the sentence level. Topic, however, is a widely used term, often with very different interpretations: what the sentence is about (Gundel & Fretheim 2004); the presupposed part of the sentence (Jackendoff 1972); the first ideational ('contentful') element in the sentence, referred to as *topical* Theme (Halliday 1967), or only those ideational elements that refer to entities (Downing 1991). For an excellent overview of the terms and their meaning, see Gómez-González (2001).

Given the impreciseness of the term 'topic', the creators of Centering Theory decided to use the term 'center' for all entities in the discourse, and 'backward-looking center' for the particular entity in the current utterance that was most salient in the previous utterance (Grosz & Sidner 1998; Joshi & Weinstein 1998). In order to determine which is the most salient entity, a ranked list of entities needs to be created, the list of forward-looking centers (Cf list).

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. 1994). In German, Strube and Hahn (1999) incorporated discourse status (Prince 1981). In Italian (Di Eugenio 1998) and Turkish (Turan 1995), empathy was also considered important. The analysis described here uses the most common template for English, represented in (6). Subjects are always ranked higher, except for impersonal and arbitrary Subjects (described below). The category Other includes mostly Adverbials.

(6)   Subject > Object(s) > Other

The starting point for ordering the Cf list in Spanish is the same as for English: Subjects are ranked higher than Objects. Since Spanish is a pro-drop language, null or zero Subjects are included in the ranking, and also ranked higher than Objects. Exceptions are arbitrary and impersonal Subjects, in sentences such as "People like it here" of "One/you like(s) it here", which are ranked lowest in the scale. Arbitrary pronouns (Jaeggli 1986; Suñer 1983) refer to person(s) that need not be defined or filled by a specific referent, or that are actually not defined in the context ("They're knocking on the door"). Suñer (1983) showed that the null pronoun in Spanish may have an arbitrary interpretation (*Dicen que va a nevar,* '(They) say that it is going to snow').

In Spanish, besides plural third person and singular second person, arbitrary pronouns are represented in impersonal sentences with the pronoun *se* (García 1975): *Se vive mejor en España* ('One lives better in Spain'). In example (7), the speaker tells the interlocutor that he sounds well (he had been sick), using an impersonal form: one can hear that you sound well. This can be considered an impersonal in middle voice, according to Mendikoetxea (1999).[7]

(7)   Ya      se  te      oye            muy  bien.
      already se  CL:2SG  hear:3SG.PRES  very  well
      'You already sound very well.'          Cf: B (speaker, *te*), IMPERSONAL (*se*)

Two other criteria are taken into account for ranking: empathy and animacy. Following Di Eugenio (1998), we take empathy with the speaker or hearer over strict word order as a ranking criterion. Empathy, as defined by Kuno (1987), "is the speaker's identification, which may vary in degree, with a person/thing that participates in the event or state that he describes in a sentence."

Empathy in Spanish is reflected in the Experiencer in psychological verbs, where the point of view taken is that of the Experiencer, regardless of whether it is the Subject or not (e.g., 'it seems to *me*', '*I* think', and similar verbs). In example (8), the clitic *me* refers to the speaker, for whom Thursday is a better date. A number of verbs in Spanish follow this pattern (*me conviene* 'it's good for me', *me viene mejor* 'it's better for me',

*se me hace que* '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,
      CL:1SG  come-3SG.PRES  better  the  Thursday
      'Thursday is better for me,'      Cf: I (*me*), IT (the meeting, null), THURSDAY

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 *se* 'to her' precedes the direct *lo* '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  CL:3SG:DAT  CL:3SG:MASC:ACC  give:3PL:PRES
           'And they gave it to her!'
                                         Cf: SISTER (*se,* 'to her'), PROGRAM (*lo,* 'it'), THEY (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 CONTINUE transition, the Subject of $U_2$ is a pronoun, or a null pronoun in languages that allow it. When the transition is a RETAIN, regular or stressed pronouns are used instead. In this section,

**Table 2.** Number of words and utterances in each corpus

|            | English | Spanish |
|------------|---------|---------|
| Words      | 11,457  | 8,694   |
| Utterances | 1,174   | 1,198   |

I show some of those tendencies for a corpus of Spanish and English conversations, and also what factors are involved when the general trend is not followed.

The study was conducted on five conversations in English and five in Spanish. The conversations were selected from the corresponding version of the CallHome corpus, distributed through the Linguistics Data Consortium. The conversations were selected to be all of similar length. They are telephone conversations between two speakers who know each other. Occasionally, more than two speakers participate, when one of the speakers passes the phone to another person in the same household.

Table 2 presents a summary of the number of words and utterances in each language. Utterance is defined in Section 3, and in general it refers to a finite clause. The rest of this section presents the results obtained for each of the two languages, and how they compare to each other.

## 5.1   Referring expressions and transitions

The first step in the analysis is to identify what transitions hold between utterances. For that purpose, Centering structures are created for each utterance: list of forward-looking centers (Cf), backward-looking center (Cb), and preferred center, the first entity in Cf (Cp). Then, we compare each pair of utterances according to Table 1. This yields four different types of transitions: CONTINUE, RETAIN, SMOOTH SHIFT, and ROUGH SHIFT.

Some utterances do not have a backward-looking center. This may be due to a number of reasons: they start the conversation; they start a new topic; or they contain no entities that can be clearly related to the previous utterance. A total of 395 utterances had an empty Cb in English, and 408 in Spanish. These utterances cannot be considered in the analysis, since Centering relies on the presence of a Cb. Some other utterances are discarded, because they contained no entities. After removing these two subsets, we came to a total of 779 utterances in English, and 790 in Spanish with a clearly defined transition. The transitions were classified into the four types, as shown in Table 3.

The results are interesting in that they confirm the prediction that certain transition types are preferred over others, in the following order: CONTINUE > RETAIN > SMOOTH SHIFT > ROUGH SHIFT. The prediction holds for both languages. Another study of different conversations in Spanish (task-oriented dialogues) showed a similar distribution of transitions (Taboada 2002). It is clear that, when the topic is not being changed (which results in an empty Cb, and is outside the results in Table 3), speakers prefer to continue talking about exactly the same entity about 65% of the time.

**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 | |
|---|---|---|---|---|---|---|---|---|
|  | English | Spanish | English | Spanish | English | Spanish | English | Spanish |
| Zero pr. | 5% | 49.5% | – | 26% | 4.5% | 49% | – | 16.7% |
| Pronoun | 80% | 9% | 54% | 6% | 67% | 6% | 58% | 7% |
| Clitic | – | 21% | – | 33% | – | 20% | – | 43% |
| NP | 12% | 16% | 24% | 18% | 23% | 19% | 17% | 27% |
| Dem. pr. | 2% | 2.5% | 4% | 2% | 4.5% | 3% | 4% | 3% |
| Other | 2% | 2% | 18% | 14% | 1% | 3% | 21% | 3% |
| n | 511 | 515 | 132 | 129 | 112 | 116 | 24 | 30 |

The next step in the analysis consists of relating the transition type to the anaphoric term used for the backward-looking center in the utterance, which can be considered the topic of that utterance. Each Cb was coded, and divided into one of the following categories: zero pronoun, pronoun, clitic, noun phrase, demonstrative pronoun, and other. At first we also coded possessive determiners, wh-words and adjectives that contain some reference (e.g., gender and number in Spanish). The numbers for each of those were so small that we decided to group them under the "Other" category. Examples (11)–(17) provide some representative instances of each type of Cb, with the Cb in question in italics. Some categories, of course, only apply to Spanish (clitic and, to a certain extent, zero pronoun). Table 4 presents the results of the analysis, and the following sections examine the distribution within each transition type.

(11)  *Zero pronoun*
   A:  a.  Elisa, ¿cómo anda?
         'Elisa, how's she?'
   B:  b.  Bien, __ está bien, está sigue laburando donde laburaba.
         'Well, (*she*) is well, she's she continues to work where she worked.'

(12)  *Pronoun*
   B:  a.  so I might I'm going to my sister's tomorrow
       b.  *she* needs a babysitter on Thursday

(13) *Clitic*

    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?
             'Yeah, what does (he) have?'

    B:  c.   *lo* hice comprar una de éstas…
             '(I) made *him* buy one of these…'

(14) *Noun phrase*

    A:  a.   and she said to ask about the food what kind of food do you have

    B:  b.   oh *the food* is fine they're they're uh

(15) *Demonstrative pronoun*

    A:  a.   So now what is carnival?

        b.   Does *that* have to do with Easter or not? No.

(16) *Other – Possessive determiner*

    B:  a.   I want to tell you

        b.   that when I got to Chicago

        c.   *my* sister Agnes was right there to meet me

(17) *Other – Possessive pronoun*[8]

    B:  a.   Sólo recibiste esas cuatro.
             'You only got those four [letters].'

    A:  b.   Y una de Marta.
             'And one from Marta.'

    B:  c.   Cinco.
             'Five.'

    A:  d.   Sí.
             'Yes.'

    B:  e.   Bueno, *mía*, mañana seguro que ya te llega por lo más tardar,
             'Well, *mine*, I'm sure you'll get tomorrow at the latest,'

## 5.2 Continue transitions

In the CONTINUE transitions, most Cbs are realized as pronouns in English (80% of the total number of Cbs in English), and as zero pronouns in Spanish (49.5%). That is to be expected: if the topic is continued from the previous utterance, the least marked anaphoric term available in the language is used – pronoun in English, zero pronoun in Spanish (Bentivoglio 1983; Givón 1983). However, the percentages are quite different: pronoun is used much more often in English than in Spanish. That is because Spanish has yet another type of anaphoric term available: clitics. These account for 21% of the CONTINUE transition Cbs. In addition, Spanish can also resort to a Subject pronoun, just as English does.

An interesting aspect in English is the presence of some zero pronouns, as in example (18). The speakers are discussing the rate for a long-distance phone call. Speaker B called somebody else, and explains that the rate, 'it' in (18a), was the medium price. Then, after A's backchannel, Speaker B refers back to the rate, but does not use a pronoun at all. Subject ellipsis in English is, although generally considered rare, quite common in certain conversational contexts (Nariyama 2004).

(18)   B:   a.   it was like the medium price
      A:   b.   yeah
      B:   c.   ___ wasn't the cheapest

Adding up zero and pronouns in English; and zero, pronoun and clitics for Spanish, we find very similar percentages: 85% versus 79.5%. The next question is: how does Spanish choose among those three types? It is obvious that the choice between pronoun and clitic is determined by syntactic function: only Objects can be expressed as clitics. But why does a speaker choose an overt Subject pronoun over a zero pronoun? The answer to that question is, in part, related to dialectal variation in Spanish (Cameron 1992; Enríquez 1984), with some dialects using pro-drop more often than others. There are also pragmatic constraints on the presence of a Subject pronoun, such as claiming the floor for an extended period of time, especially in the case of first and second person pronouns (Davidson 1996). Stewart (1999) proposes that the use of the first person singular pronoun is a politeness resource, which helps contrast the speaker with other individuals or groups. Luján (1999) also points out the contrastive character of first and second person pronouns. In example (19), the speaker seems to be emphasizing his role as an actor in the narrative he is constructing. He is discussing how he came to realize that he had a rash. In (19a), he is himself the backward-looking center. In (19b), he is also the backward-looking center. However, in (19b), he uses *yo* ('I') to describe his own thought processes. This utterance could be translated as 'I thought, maybe it's something [serious]'.

(19)   a.   de repente me salió un granito en la ingle,
          'all of a sudden I got a zit in my groin,'
     b.   yo dije, uy, yayay, no vaya a ser algo –
          'I said, uh-oh, maybe it's something…'

Third person pronouns are also expressed in Spanish, even when they are clear from context, for emphasis. In (20), the speakers are discussing B's friend, who applied for a scholarship and got it. The friend is the Cb in (20a), because the Subject is impersonal, and therefore lower in the Cf list. In (20c), the friend is referred to with a pronoun, because it is emphasized.

(20)   B.   a.   y se la dieron
           'and they gave it [=the scholarship] to her'
     A:   b.   Ah, ¿sí? {laughter}
           'Really?'

  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 CONTINUE 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), JOHNNY (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 *we* ('we camped there', 'we rented a bike'). Speaker B mentions the name *Johnny*, 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 JOHNNY 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 RETAIN transitions

The percentage of pronoun forms (whether zero, pronoun, or clitic) in the RETAIN transitions decreases considerably, as compared to the CONTINUE transitions. In a RE-TAIN 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 PUBLIC BEACH to I, the speaker. Since PUBLIC BEACH 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: PUBLIC BEACH – Cb: PUBLIC BEACH

    B: b. oh really
    A: c. I mean I didn't I didn't know where it was

              Cf: I, PUBLIC BEACH – Cb: PUBLIC BEACH

The Cb may, in RETAIN transitions, still be expressed through a pronoun. According to Di Eugenio (1998), this is possible especially in cases where the pronominalized

(Subject) Cb unequivocally refers to an entity other than the previous Cb. This unequivocal reference can be achieved through syntactic features such as person and number. In the previous example, person differentiates the pronominalized Subject *I* from the pronominalized Cb *it*.

In cases where the reference is not clear, speakers use other anaphoric terms to encode the Cb of a RETAIN transition, which accounts for the decrease in pronoun forms from the CONTINUE transition. In example (23), there are three different entities: SHE, a third person female encoded in (23a) as *her*, MOTHER, and COUSIN. In (23b), the focus shifts from HER COUSIN, the Subject in (23a), to SHE, the Subject in the new utterance. Because at least two of these entities (SHE, HER MOTHER), and maybe even HER COUSIN are female, the use of a pronoun to refer to the backward-looking center HER MOTHER would have been infelicitous. The speaker avoids confusion by using a full noun phrase.

(23)  a.  First, you know, her mother her cousin is a manager of a travel agency.
      b.  And she invited her mother and her fr- uh and the cousin to come on out.

Some of the uses of anaphoric terms other than pronouns can be related to spoken language phenomena, such as holding the turn. In example (24c), which shows a RETAIN transition in relation to (24b), Speaker A repeats *mi hermana* ('my sister'), the Cb, as a full noun phrase. It seems that he is holding the turn while he thinks about what he can tell his interlocutor. Notice that between (24a) and (24b) there is a CONTINUE transition, with the Cb MY SISTER also encoded as an NP.[9]

(24)  B:  a.  ¿Y tu hermana?
              'And your sister?'                    Cf: A's SISTER – Cb: Ø
      A:  b.  Mi hermana está bie-
              'My sister is fine'                   Cf: A's SISTER – Cb: A's SISTER
          c.  No, hombre, qué te platico de mi hermana,
              'Well, what can (I) tell you about my sister,'
                  Cf: A (null), B (*te*), A's SISTER (*mi hermana*) – Cb: A's SISTER

## 5.4 SMOOTH SHIFT transitions

SMOOTH SHIFTS represent the cases in which speakers are preparing to shift topics: the previous topic is still present in the utterance, but a new entity has gained prominence, with the Cb of the current utterance being different from the Cb of the previous one. However, in our data, the types of pronouns used do not differ very much from those in CONTINUE transitions. The three pronoun types (zero, pronoun and clitic, where applicable) still account for the majority of Cb expressions. The only significant change, for both languages, is that few Cbs fall under the "Other" category. This could be because it is necessary to encode the Cb with a form that makes the reference unambiguous. Some of the "Other" forms, such as demonstrative pronouns or adverbials might make this difficult.

Example (25) shows a case of RETAIN with a clitic (*me,* 'me'). In this case, the reference is clear because of the switch between first and third person ('I told her', 'She told me').

(25)  a.  Yo le dije, sabes qué, maestra, estoy enfermísimo.
          'I told her, (you) know what, teacher, (I)'m really sick.'
      b.  Me dijo, no me bronca.
          '(She) told me, (it) doesn't matter to me.'

In example (26), Speaker B is discussing her father's dizziness. The entities in (26a) are IT (the dizziness) and HEAT. Speaker A picks up on the least prominent of those utterances, HEAT, and elaborates on it. The shift from DIZZINESS to HEAT requires that Speaker A use a full noun phrase.

(26)  B:  a.  and maybe it was because of the was the heat or something
                    Cf: DIZZINESS (*it*), HEAT (*the heat*) – Cb: DIZZINESS
      A:  b.  maybe the heat it's so God damn hot here...
                    Cf: HEAT, HERE – Cb: HEAT

## 5.5  ROUGH SHIFT transitions

Finally, ROUGH SHIFTS represent a change in the topic, while still preserving the same entities from one utterance to the next (a radical change of topic results in an empty Cb; those utterances were not considered in the analysis). Again, we see that the majority of Cbs are encoded using one of the three pronoun categories, with English using exclusively the full pronoun, and Spanish again distributing most of the Cbs into either zero pronouns or clitics, depending on their grammatical function. The most significant difference between the two languages in ROUGH SHIFTS is under "Other", used much more often in English. The numbers of ROUGH SHIFTS, are, however, smaller, which makes generalizations more difficult.

In some cases, a pronoun is the only possibility, because the pronouns are used to refer to the speakers themselves. In (27c), the Cb is the speaker, the highest-ranked entity from (27b). But since in (27c) the interlocutor is part of the Cf list (the addressee of the imperative), the speaker is no longer the highest-ranked member of the Cf list, which results in a rough shift. The pronoun *me* is enough to understand the reference. Furthermore, no other form could have been used. It is also interesting to point out that this example conforms to Rule 1 of Centering Theory: the Cb of (27b) is a pronoun, and no other pronoun is used in the example.

(27)  a.  there's no nice blessed mother over here {laugh}
                    Cf: BLESSED MOTHER, HERE
      b.  and I would really like it for my bedroom
                    Cf: B (*I*), BLESSED MOTHER (*it*), BEDROOM – Cb: BLESSED MOTHER

c. just send me the blessed mother not the cross
>    Cf: A (imperative), B (*me*), ʙʟᴇssᴇᴅ ᴍᴏᴛʜᴇʀ, ᴄʀᴏss – Cb: B

Similarly, in (28), the only possible form of the Cb is the noun phrase *el treinta y uno* ('the thirty-first', New Year's Eve), which is part of a prepositional phrase. Speaker B needs to make clear which one of the dates she is picking, and the repetition of the noun phrase seems necessary.

(28)  A:  a.  Y no tienes todavía programa ni para el veinticuatro ni para el treinta y uno.
>        'And you have no plans yet for the twenty-fourth or for the thirty-first.'
>    B:  b.  Para el treinta y uno estamos viendo
>        'For the thirty-first we're checking…'

## 6. Conclusions

Speakers manage topics and referents in conversation with incredible ease. The purpose of this paper has been to explore some of the management strategies used by speakers in English and Spanish. Based on a corpus analysis of five English and five Spanish conversations, I have discussed some of the factors involved in the choice of anaphoric term for a given referent.

The framework of the analysis is Centering Theory, which provides clear definitions of 'topic' progression, and the relations between pairs of utterances in discourse. Using Centering, the analysis shows some tendencies in what anaphoric term is chosen for the backward-looking center (Cb) of a given utterance, what could loosely be characterized as the topic of the utterance. The choice of anaphoric term for the Cb is related to whether the topic is continued, slightly changed or abruptly changed from the utterance immediately before.

The results show that speakers tend to use pronominal forms when the topic is continued (ᴄᴏɴᴛɪɴᴜᴇ transitions), and fewer pronominal forms when there is a slight change in topic (ʀᴇᴛᴀɪɴ transitions). This was not surprising at all; more interesting are the cases where speakers do not choose a pronoun (although one could be chosen without creating ambiguity); or where speakers use a pronoun instead of a more explicit form. In most cases, we can relate the choices against expectation to spoken language phenomena: speakers use a 'heavier' form than expected (e.g., full noun phrase instead of pronoun) when they want to emphasize the referent. They also tend to use heavier forms across turns, to signal familiarity with the referent, or that the referent is clearly established for both interlocutors. On the other hand, they use less explicit forms (e.g., pronoun when a noun would have been expected) if the pronoun carries enough grammatical information to make the referent clear.

Future directions of this research include a more thorough examination of proper name use in spoken conversation, and clitic doubling in Spanish (Suñer 1988). We

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.

**1.** I use 'topic' in quotes, because in this paper it refers exclusively to the backward-looking center, which is a well-defined term (see Section 2).

**2.** The order of the two Objects is variable in the literature. In the analyses carried out so far, we have ranked Indirect before Direct. This also corresponds to the surface order of the sentence in the example.

**3.** Small capitals indicate that the list contains entities, not their linguistic realization. In Cornish's (1999) terms, it contains referents, not antecedents or anaphors.

**4.** Centering transitions are one possible explanation for coherence. A text can be coherent without repeating or referring to the same entities (Brown & Yule 1983; Poesio et al. 2000).

**5.** Similar to the T-unit used in developmental studies and in text analysis (Hunt 1977).

**6.** A higher number of empty backward-looking centers is undesirable, since it indicates that Centering structures found few connections between consecutive pairs of utterances. It is assumed that most utterances will have a connection with the preceding speech.

**7.** Spanish examples are glossed only when word order or grammatical information are considered relevant. Otherwise, they are simply translated. Abbreviations used in the glosses: 1/2/3 ('first/second/third person'); CL ('clitic'); NOM ('nominative'); ACC ('accusative'); DAT ('dative'); SG ('singular'); PL ('plural'); FEM ('feminine'); MASC ('masculine'); PRES ('present'); PRET ('preterite').

**8.** In example (17e), *mía* ('mine') is the backward-looking center. It refers to *cinco* ('five [letters]') in (17c). (17d) was not part of the analysis, because it is a backchannel that contains no entities.

**9.** Another interpretation of this example would break the noun phrase *mi hermana* ('my sister'), and extract two entities from it: I (*mi*) and SISTER (*hermana*). I have taken that approach elsewhere, but will ignore that distinction for illustration purposes in this example.

## References

Ariel, Mira (1990). *Accessing Noun Phrase Antecedents*. London: Routledge.

Bentivoglio, Paola (1983). "Topic continuity and discontinuity in discourse: A study of spoken Latin-American Spanish". In T. Givón (Ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study* (pp. 255–311). Amsterdam and Philadelphia: John Benjamins.

Brennan, Susan E. (1995). "Centering attention in discourse". *Language and Cognitive Processes, 10*, 137–167.

Brennan, Susan E., Marilyn W. Friedman, & Carl J. Pollard (1987). "A Centering approach to pronouns". In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics* (*ACL-87*) (pp. 155–162). Stanford, CA, USA.

Brown, Gillian & George Yule (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.

Byron, Donna K. & Amanda Stent (1998). "A preliminary model of Centering in dialog". In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics* (*ACL-98*) (pp. 1475–1477). Montréal, Canada.

Cameron, Richard (1992). "Pronominal and Null Subject Variation in Spanish: Constraints, Dialects, and Functional Compensation." Ph.D. dissertation. University of Pennsylvania.

Chafe, Wallace (1976). "Givenness, contrastiveness, definiteness, subjects, topics, and point of view". In C. Li (Ed.), *Subjects and Topics* (pp. 26–56). New York: Academic Press.

Cornish, Francis (1999). *Anaphora, Discourse, and Understanding: Evidence from English and French*. Oxford: Clarendon Press.

Daneš, František (1974). "Functional Sentence Perspective and the organization of the text". In F. Daneš (Ed.), *Papers on Functional Sentence Perspective* (pp. 106–128). Prague: Academia.

Davidson, Brad (1996). "'Pragmatic weight' and Spanish subject pronouns: The pragmatic and discourse uses of 'tú' and 'yo' in spoken Madrid Spanish". *Journal of Pragmatics, 26*, 543–565.

Di Eugenio, Barbara (1998). "Centering in Italian". In M. A. Walker, A. K. Joshi, & E. F. Prince (Eds.), *Centering Theory in Discourse* (pp. 115–137). Oxford: Clarendon Press

Downing, Angela (1991). "An alternative approach to Theme: A systemic-functional perspective". *Word, 42*, 119–143.

Downing, Pamela A. (1996). "Proper names as a referential option in English conversation". In B. A. Fox (Ed.), *Studies in Anaphora* (pp. 95–143). Amsterdam and Philadelphia: John Benjamins.

Eckert, Miriam & Michael Strube (1999). "Resolving discourse deictic anaphora in dialogues". In *Proceedings of the 9th Conference of the European Chapter of the Association for Compu–tational Linguistics* (*EACL-99*) (pp. 37–44). Bergen, Norway.

Enríquez, Emilia (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas.

García, Erica (1975). *The Role of Theory in Linguistic Analysis: The Spanish Pronoun System*. Amsterdam: North-Holland.

Givón, Talmy (1983). "Topic continuity in discourse: An introduction". In T. Givón (Ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study* (pp. 1–41). Amsterdam and Philadelphia: John Benjamins.

Gómez-González, María de los Ángeles (2001). *The Theme-Topic Interface: Evidence from English*. Amsterdam and Philadelphia: John Benjamins.

Gordon, Peter C., Barbara J. Grosz, & Laura A. Gilliom (1993). "Pronouns, names, and the Centering of attention in discourse". *Cognitive Science, 17*, 311–347.

Gordon, Peter C., Randall Hendrick, Kerry Ledoux, & Chin Lung Yang (1999). "Processing of reference and the structure of language: An analysis of complex noun phrases". *Language and Cognitive Processes, 14*, 353–379.

Grosz, Barbara J., Aravind K. Joshi, & Scott Weinstein (1995). "Centering: A framework for modelling the local coherence of discourse". *Computational Linguistics, 21*, 203–225.

Grosz, Barbara J. & Candace L. Sidner (1986). "Attention, intentions, and the structure of discourse". *Computational Linguistics, 12*, 175–204.

Grosz, Barbara J. & Candace L. Sidner (1998). "Lost intuitions and forgotten intentions". In A. Walker, A. K. Joshi, & E. F. Prince (Eds.), M*Centering Theory in Discourse* (pp. 39–51). Oxford: Clarendon Press.

Gundel, Jeanette K., Nancy Hedberg, & Ron Zacharski (1993). "Cognitive status and the form of the referring expressions in discourse". *Language, 69*, 274–307.

Gundel, Jeanette K. & Thorstein Fretheim (2004). "Topic and focus". In L. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 175–196). Oxford: Blackwell.

Hadic Zabala, Loreley & Maite Taboada (2004). "Centering Theory in Spanish: Coding Manual". Manuscript. Simon Fraser University.

Halliday, Michael A. K. (1967). "Notes on Transitivity and Theme in English. Part II". *Journal of Linguistics, 3*, 199–244.

Hudson-D'Zmura, Susan & Michael K. Tanenhaus (1998). "Assigning antecedents to ambiguous pronouns: The role of the center of attention as the default assignment". In M. A. Walker, A. K. Joshi, & E. F. Prince (Eds.), *Centering Theory in Discourse* (pp. 199–226). Oxford: Clarendon Press.

Hunt, Kellogg W. (1977). "Early blooming and late blooming syntactic structures". In C. R. Cooper & L. Odell (Eds.), *Evaluating Writing: Describing, Measuring and Judging* (pp. 91–104). Urbana, IL: NCTE.

Hurewitz, Felicia (1998). "A quantitative look at discourse coherence". In M. A. Walker, A. K. Joshi, & E. F. Prince (Eds.), *Centering Theory in Discourse* (pp. 273–291). Oxford: Clarendon Press.

Iida, Masayo (1998). "Discourse coherence and shifting centers in Japanese texts". In M. A. Walker, A. K. Joshi, & E. F. Prince (Eds.), *Centering Theory in Discourse* (pp. 161–180). Oxford: Clarendon Press.

Jackendoff, Ray (1972). *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.

Jaeggli, Osvaldo A. (1986). "Arbitrary plural pronominals". *Natural Language and Linguistic Theory, 4*, 43–76.

Joshi, Aravind K. & Scott Weinstein (1998). "Formal systems for complexity and control of inference: A reprise and some hints". In M. A. Walker, A. K. Joshi, & E. F. Prince (Eds.), *Centering Theory in Discourse* (pp. 31–38). Oxford: Clarendon Press.

Kameyama, Megumi (1985). "Zero Anaphora: The Case of Japanese". Ph.D. dissertation. Stanford University.

Kameyama, Megumi (1998). "Intrasentential Centering: A case study". In M. A. Walker, A. K. Joshi, & E. F. Prince (Eds.), *Centering Theory in Discourse* (pp. 89–112). Oxford: Clarendon Press.

Kuno, Susumu (1972). "Functional Sentence Perspective: A case study from Japanese and English". *Linguistic Inquiry, 3*, 269–320.

Kuno, Susumu (1987). *Functional Syntax: Anaphora, Discourse and Empathy*. Chicago: University of Chicago Press.

Lambrecht, Knud (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press.

Luján, Marta (1999). "Expresión y omisión del pronombre personal". In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (pp. 1275–1315). Madrid: Espasa.

Mendikoetxea, Amaya (1999). "Construcciones con *se*: medias, pasivas e impersonales". In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (pp. 1631–1722). Madrid: Espasa.

Miltsakaki, Eleni (2001). "Centering in Greek". Paper presented at the *15th Symposium on Theoretical and Applied Linguistics*. Thessaloniki, Greece.

Miltsakaki, Eleni (2002). "Toward an aposynthesis of topic continuity and intrasentential anaphora". *Computational Linguistics, 28*, 319–355.

Nariyama, Shigeko (2004). "Subject ellipsis in English". *Journal of Pragmatics, 36*, 237–264.

Poesio, Massimo (2000). "The GNOME annotation scheme manual". Technical Report. University of Edinburgh. (Available from: http://www.cogsci.ed.ac.uk/~gnome/)

Poesio, Massimo, Hua Cheng, Renate Henschel, Janet Hitzeman, Roger Kibble, & Rosemary Stevenson (2000). "Specifying the parameters of Centering Theory: A corpus-based evaluation using text from application-oriented domains". In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics* (*ACL-2000*) (pp. 400–407). Hong Kong.

Prasad, Rashmi & Michael Strube (2000). "Discourse salience and pronoun resolution in Hindi". *Penn Working Papers in Linguistics, 6*, 189–208.

Prince, Ellen F. (1981). "Towards a taxonomy of given-new information". In P. Cole (Ed.), *Radical Pragmatics* (pp. 223–255). New York: Academic Press.

Schiffrin, Deborah (1994). *Approaches to Discourse*. Malden, MA: Blackwell.

Stewart, Miranda (1999). "Hedging your bets: The use of *yo* in face-to-face interaction". *Web Journal of Modern Language Linguistics, 4–5*, http://wjmll.ncl.ac.uk/issue04-05/stewart.htm

Strube, Michael & Udo Hahn (1996). "Functional Centering". In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics* (*ACL-96*) (pp. 270–277). Santa Cruz, CA, USA.

Strube, Michael & Udo Hahn (1999). "Functional Centering – Grounding referential coherence in information structure". *Computational Linguistics, 25*, 309–344.

Suñer, Margarita (1983). "pro$_{arb}$". *Linguistic Inquiry, 14*, 188–191.

Suñer, Margarita (1988). "The role of agreement in clitic-doubled constructions". *Natural Language and Linguistic Theory, 6*, 391–434.

Taboada, Maite (2002). "Centering and pronominal reference: In dialogue, in Spanish". In J. Bos, M. E. Foster, & C. Matheson (Eds.), *Proceedings 6th Workshop on the Semantics and Pragmatics of Dialog, EDILOG* (pp. 177–184).

Taboada, Maite (forthcoming). "Reference, centers and transitions in spoken Spanish". In J. Gundel & N. Hedberg (Eds.), *Reference and Reference Processing*. Oxford: Oxford University Press.

Turan, Ümit Deniz (1995). "Null vs. Overt Subjects in Turkish Discourse: A Centering Analysis". Ph.D. dissertation. University of Pennsylvania.

Vallduví, Enric (1990). "The Informational Component". Ph.D. dissertation. University of Pennsylvania.

Vallduví, Enric & Elisabet Engdahl (1996). "The linguistic realization of information packaging". *Linguistics, 34*, 459–519.

Walker, Marilyn A., Masayo Iida, & Sharon Cote (1994). "Japanese discourse and the process of Centering". *Computational Linguistics, 20*, 193–232.

Walker, Marilyn A., Aravind K. Joshi, & Ellen F. Prince (1998). "Centering in naturally occurring discourse: An overview". In M. A. Walker, A. K. Joshi, & E. F. Prince (Eds.), *Centering Theory in Discourse* (pp. 1–28). Oxford: Clarendon Press.

Yngve, Victor H. (1970). "On getting a word in edgewise". In *Papers from the Sixth Regional Meeting of the Chicago Linguistics Society* (pp. 567–577). Chicago: University of Chicago.