Reference, Centers and Transitions in Spoken Spanish[*]

Maite Taboada

Simon Fraser University

mtaboada@sfu.ca

Abstract

The goal of this paper is to examine the relationship between Centering transitions (Grosz et al., 1995) and choice of referring expression. For that purpose, Centering analyses were carried out in two different corpora of spoken Spanish. The corpus analysis confirms reports in previous literature about what is the typical choice of referring expression. In some cases, however, the referring expression chosen violates expectation, or does not follow what other researchers have found (e.g., a proper name is used when a pronoun is expected). In those cases, the most likely explanation is that other constraints related to spoken language are at play (turn-taking and grounding).

1. Introduction

The question that much of the research on anaphora attempts to answer is: how does a speaker choose which referring expression to use? One assumption is that the speaker uses the referring expression that conveys the exact amount of information that the hearer will need in order to interpret the current utterance correctly. Given a possible choice between *he, this man, the man,* and *John*, it is plausible that a speaker will choose one that will help the hearer link to the intended referent with the minimum amount of effort. If the conversation has been about John throughout, with no other male referent intervening, *he* is probably the most common choice. If the speaker uses *John* instead, she might indicate that the hearer is to pay attention to the referent, or that a new John has been introduced in the conversation. Any explanation needs to not only account for the most typical realization (i.e., the expected realization), but also explain what factors are involved when the choice is contrary to expectation. Bolinger formulates the question in the following terms:

> "At *X* location, what reason might the speaker have for using a word that is leaner in semantic content rather than one that is fuller, or vice versa?" Usually this means "Why use a pronoun?" or "Why repeat the noun?" (Bolinger, 1979: 290)

Different explanations have been proposed to account for how the choices are made, and for the effects of such choices, such as Gundel et al.'s (1993) Givenness Hierarchy or Ariel's (1996) accessibility marking scale. In these, the form of the referring expression is linked to the salience of the referent. Other explanations emphasize the importance of first mention (Carreiras et al., 1995; Gernsbacher and Hargreaves, 1988), or syntactic organization (Gordon et al., 1999).

In this paper, I explore a different way to explain the form of a referring expression, by applying Centering Theory (Grosz et al., 1995). Centering Theory is a theory of local focus in discourse that proposes different transition types between any pair of utterances. Those transitions are based on salience, but also on the expectations that the hearer might have about the focus of the next utterance. Researchers within Centering Theory have already proposed that there is a relation between the form of a referring expression in a given utterance and the transition linking that utterance to the previous one, or that Centering structures guide the interpretation of pronouns in discourse (Brennan, 1995; Di Eugenio, 1998; Gordon et al., 1993; Hudson-D'Zmura and Tanenhaus, 1998; Roberts, 1998; Walker, 1998). I extend that research by applying Centering to Spanish spoken discourse.

It should be obvious that transition type is not the only factor involved: Centering proposes four transition types; most languages number more than four choices in their repertoire of referring expressions, meaning that more than four referring forms are possible for a given entity. For example, Gundel et al. (1993) propose six cognitive statuses and at least seven different referring expressions in English that denote them. That means that other factors must be at play in the choice. The paper also explores some of those factors.

The study was carried out on two corpora of spoken Spanish. The first one, the Interactive Systems Lab corpus, is a collection of task-oriented conversations between two speakers. The second one is the CallHome corpus, a set of telephone conversations between relatives or friends. A total of fourteen conversations from the two corpora were annotated according to Centering theory.

The paper is structured as follows: Section 2 will briefly introduce Centering Theory; Section 3 describes its application to spoken discourse, in particular as regards to segmentation. Section 4 explains the process of constructing the list of entities, the Cf list. The results of the corpus analysis are presented and discussed in Section 5, with Section 6 providing conclusions.

2. Centering Theory

Centering (Grosz et al., 1995; Walker et al., 1998) was developed within a theory of discourse structure (Grosz and Sidner, 1986) that considers the interaction between (i) the intentions, or purposes, of the discourse and 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 4). 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) illustrates these concepts[1]. Let us assume that the utterances in the example constitute a discourse segment. In the first utterance, (1a), there are two centers: *Harry* and *snort*. (1a) does not have a backward-looking center (the center is empty), because this is the first utterance in the segment. In (1b), two new centers appear: *the Dursleys* and *their son, Dudley*. The lists include centers ranked according to two main criteria: grammatical function and linear order. (Ranking will be further discussed in Section 4.) The Cf list for (1b) is: DURSLEYS, DUDLEY[2]. The preferred center in that utterance is the highest-ranked member of the Cf list, i.e., DURSLEYS. The Cb of (1b) is empty, since there are no common entities between (1a) and (1b). In (1c), a few more entities are presented, and they could be ranked in a number of ways. To shorten the discussion at this point, I will rank them in linear order, left-to-right. In any event, the most important entities seem to be the Subject, which is the same as in (1b), DURSLEYS; and DUDLEY, realized by in the possessive adjective *his* (twice). The Cp is DURSLEYS, since it is the highest-ranked member of the Cf list, and the Cb is also DURSLEYS, because it is the highest-ranked member of (1b) repeated in (1c). The new utterance, (1d), reintroduces Harry to the discourse, and links to (1c) through DUDLEY, which is the Cb in (1d).

(1) a. Harry suppressed a snort with difficulty.

    b. The Dursleys really were astonishingly stupid about their son, Dudley.

    c. They had swallowed all his dim-witted lies about having tea with a different member of his gang every night of the summer holidays.

    d. Harry knew perfectly well that Dudley had not been to tea anywhere;

    e. he and his gang spent every evening vandalising the play park, [...]

In (2) we see the Cf, Cp and Cb for each of the utterances in the segment:

(2) a. Cf: HARRY, SNORT

      Cp: HARRY – Cb: Ø

    b. Cf: DURSLEYS, DUDLEY

      Cp: DURSLEYS – Cb: Ø

    c. Cf: DURSLEYS, DUDLEY, LIES, TEA, MEMBER, GANG, NIGHT, HOLIDAYS

      Cp: DURSLEYS – Cb: DURSLEYS

    d. Cf: HARRY, DUDLEY, TEA

      Cp: HARRY – Cb: DUDLEY

    e. Cf: DUDLEY, GANG, EVENING, PARK

      Cp: DUDLEY – Cb: DUDLEY

In addition to the different types of centers, Centering proposes transition types, based on the relationship between the backward-looking centers of any given pair of utterances, and the

---

[1] From J.K. Rowling (2003) *Harry Potter and the Order of the Phoenix.* Vancouver: Raincoast Books (p. 8).

[2] Small capitals indicate that the list contains entities, not their linguistic realization. The reference to Dudley is conveyed by two different referring expressions: *their son* and *Dudley*.

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_i and Cp_i refer to the centers in the current utterance. Cb_{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, in addition, 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 a conversation can progress: from how an utterance refers to a previous topic, the Cb_{i-1}, and it is still concerned with that topic, the Cp_i, in a CONTINUE, to how it can be not linked at all to the previous topic, in a ROUGH SHIFT. Transitions are one explanation[3] 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 no transition (or a zero-Cb transition). The transition between (1a) and (1b) is also zero. Between (1b) and (1c) there is a CONTINUE transition, because the Cb of (1b) is empty, and the Cp and Cb of (1c) are the same, DURSLEYS[4]. Utterance (1d) has a different Cb from (1c), and it also shows different Cb and Cp, producing then a ROUGH SHIFT in the transition between (1c) and (1d). Finally, (1e) and (1d) are linked by a RETAIN transition.

	$Cb_i=Cb_{i-1}$ or $Cb_{i-1}=\emptyset$	$Cb_i\neq Cb_{i-1}$
$Cb_i=Cp_i$	CONTINUE	SMOOTH SHIFT
$Cb_i\neq Cp_i$	RETAIN	ROUGH SHIFT

Table 1. Transition types.

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, ..., 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

---

[3] Centering transitions are just one explanation for coherence. A text can be coherent without repeating or referring to the same entities (Brown and Yule, 1983: 195-199; Poesio et al., 2000).

[4] Other proposals suggest that transitions for utterances after an empty Cb should be different: if Cb_i is not empty, but Cb_{i-1} is, the transition is a CENTER ESTABLISHMENT; if Cb_i is empty and it follows an also empty Cb_{i-1}, the transition is NULL. It is only when Cb_i is empty, and Cb_{i-1} is not that we have a ZERO transition (Kameyama, 1986; Poesio et al., 2004).

expression. In (1c), the backward-looking center, DURSLEYS, is realized as a pronoun, following Rule 1, since other pronouns are also present in the utterance (*his* to refer to DUDLEY).

Relationships have been established between the transition type between a pair of utterances, and the type of referring expression chosen to realize entities in the second utterance in the pair. Di Eugenio (1998) found that CONTINUE transitions, because they keep the same center, often encode the subject as a zero pronoun in Italian. Shifts (smooth or rough) result in less pronominalization. We will see that these relationships are quite complex, and different factors come into play in the choice of referring expression.

3. Centering and spoken language

The Centering framework has been applied to both constructed examples and naturally occurring discourse, but not widely to spontaneous conversation. There are a number of issues involved in such application, namely the segmentation into Centering units (utterances), the presence of false starts and backchannels, linearity and overlap, and the presence of first and second person pronouns. I discuss each one of those in this section.

The approach taken here to apply Centering to spoken dialogue owes much to the work done by Byron and Stent (1998). They report experiments on different variations of segmentation, false starts, inclusion of first and second person pronouns, and linearity. The model for dialogue adopted here is Byron and Stent's *Model 1*, that is, a model where both first and second person pronouns are included in the Cf list. In addition, utterances are consecutive: in the search for Cb_n, only Cf_{n-1} is searched, whether it was produced by the same speaker or not. Byron and Stent (1998) found that this model performed better than models that discarded first and second person pronouns, and models that considered previous or current speaker's previous utterance[5].

3.1 Utterance segmentation

The first step in a Centering analysis involves deciding on the minimal units of analysis, commonly referred to as 'utterances'. The notions of discourse segment and utterance are very important: Centering predicts the behaviour of entities within a discourse segment; centers are established with respect to the utterance. In this paper, I use the term 'utterance' or 'segment' to refer to the units of analysis in Centering Theory. In other applications, 'segment' or 'discourse segment' refers to the broad parts into which a discourse can be divided (e.g., introduction, thesis statement), or to discourse segments that achieve a purpose each (Grosz and Sidner, 1986). I am not concerned with those higher-level discourse segments here, but only with minimal units of analysis, typically interpreted to be either entire sentences or finite clauses. These concerns are general to Centering applications, but even more pressing when dealing with spoken language, where the notion of sentence is more difficult to instantiate. That is why, in spoken language, traditional notions of clause and sentence are abandoned in favour of the idea of an utterance (Schiffrin, 1994).

---

[5] Their performance measures were based on (i) number of zero Cbs, (ii) whether the Cb that Centering found corresponded with a loose notion of sentence topic, and (iii) number of cheap vs. expensive transitions. The cheap/expensive distinction refers to inference load on the hearer (Strube and Hahn, 1999), according to whether Cp_{n-1}, expected to be Cb_n, is actually realized as such.

In general, an utterance is an intonation unit. In the corpora studied, utterances are already marked in the transcripts. For the ISL corpus, an utterance is defined as an intonation unit marked by either a period or a question mark. Note that a comma does not always define an utterance. In Example (3), the period after *Miriam* indicates falling intonation, as in the end of a sentence. There are, therefore, two Centering units in (3)[6].

(3) a. Miriam.

  'Miriam.'

  b. yo creo que /uh/ no nos va a alcanzar el tiempo.

  'I believe that, uh, we won't have enough time.'

In the CallHome corpus, utterances, at the first level of granularity, are equivalent to dialogue acts, which were assigned to the Spanish CallHome corpus (Levin et al., 1999). In this corpus, the speech act was more important than intonation when it came to segmenting speech into utterances. The following example was segmented into two dialogue acts, which also correspond to two tensed clauses.

(4) a. Se supone que hay mucho ganado,

  'Supposedly there are a lot of animals,'

  b. pero yo no vi nada.

  'but I didn't see any.'

Pauses also indicate a new segment, whether a segment was introduced already in the transcripts or not. Example (5) was one unit, but since a pause exists after *de él*, the second part was considered to be a new Centering unit.

(5) a. claro, pero, o sea, él, según él, soy el socio de él [pause]

  'right, but, I mean, he, according to him, (I) am his partner'

  b. según él, ¿no es cierto?

  'according to him, right?'

Segmentation into utterances has been a topic of study in the Centering literature. In the analysis, I have followed Kameyama's (1998) proposals for intra-sentential Centering. They consist of separating any tensed coordinate or subordinate clauses from their matrix, and of including report complements and reported speech together with the reporting units. Tenseless subordinate clauses are part of the matrix clause[7]. In addition to the segmentation already in the corpora (utterances and dialogue acts), complex clauses are broken up according to Kameyama's rules. Tensed adjuncts are separated from the main clause, as in Example (6).

(6) a. No compro nada, no nada, nada

  '(I) don't buy anything, nothing, nothing'

---

[6] Spanish examples are glossed word-by-word only when the gloss provides information considered relevant. In all other cases, they are translated as close to the original as possible, which may sometimes make them sound awkward. Parentheses around a pronoun in the translation indicate that it is null in Spanish. Slashes (*/eh/*) indicate filled pauses or backchannels. Angle brackets (*<de>*) indicate false starts.

[7] For a more detailed explanation of the segmentation, see Hadic Zabala and Taboada (2004) and Taboada and Hadic Zabala (2005).

b. porque quiero irme a ver a mi hermana.

'because (I) want to go see my sister.'

Kameyama (1998) considers reported speech a hierarchical unit, embedded with the reporting unit, and I followed that approach. That is, in cases where reported speech appears, the reported unit is processed, and Centering structures are created within it. But once it has been processed, the next unit looks back to the reporting unit for antecedents, and for Cb comparison purposes. I also included relative clauses together with their antecedent NP, i.e., relative clauses were treated as embedded. Poesio et al. (Poesio et al., 2000; 2004) report that this produces fewer violations of Centering constraints (specifically, of Constraint 1, that all utterances of a segment, except the first one, have one Cb).

The final issue in segmentation was the speech addressed to a third party. In CallHome conversations, which are on the telephone, one of the interlocutors sometimes directs speech to another person on his or her side of the line. This was recorded, and quite likely audible to the other interlocutor. I considered speech directed to a third party as a separate Centering unit, and included it in the Centering analysis, because entities mentioned in the speech to the third party often appear in the conversation between the main interlocutors. We can see an illustration in (7). The speakers, A and B, are debating how long they have been on the phone (7a and 7b). Speaker B then asks somebody else (*mamá*), and reports back the answer. The vocative *mamá* is included in the Cf list of (7c)[8]. A Centering analysis including (7c) shows that speech directed to a third party must be included in the analysis since it contains the antecedent for the null pronoun in (7d), which is speech directed at A, and as a consequence part of the main conversation. Without (7c), the transition between (7b) and (7d) is a zero transition (no Cb).

(7) A: a. ¿Te late que como quince?

'Does fifteen (minutes) sound about right?

B: b. Pues no sé yo.

'Well, I don't know.'

c. llevamos como quince minutos, mamá?

'Have (we) been (talking) for about fifteen minutes, Mom?'

d. dice que más o menos.

'(She) says that more or less.'

The segmentation was performed by two annotators separately. We first segmented one CallHome and four ISL conversations as training, compared the results and refined the coding manual (Hadic Zabala and Taboada, 2004). Then an evaluation was performed, segmenting four additional CallHome conversations, which amounted to 895 segments in the final agreement. The disagreement in those 895 segments was 18.7% of the total. This included any instance of disagreement (two instead of one segments, or vice versa, or disagreements in the inclusion of segments for the analysis). The high disagreement rate is due to problems in interpreting spoken data (boundaries are not clear), deciding on whether to include inferables (if an utterance contains

---

[8] I believe vocatives should be part of the Cf list (see Lambrecht, 1994 about vocatives being topics, and therefore referential), but I am not sure where they belong in the Cf ranking. The current coding includes them in the highest position, following Lambrecht's (1994) suggestion that they are topics.

no entities, it is not considered a unit for the analysis), and, to a lesser extent, also due to human error. Current efforts are directed toward making the coding manual more transparent, and devising a training process, which might include segmenting on-line, without looking ahead, as Brennan (1995) suggests.

3.2 False starts and backchannels

An utterance may not be complete syntactically, but still include referential information that affects the rest of the discourse. Some of these incomplete utterances are referred to as false starts. In the analysis, I considered false starts that included some referential information, whether the utterance was complete or not, which was also the approach followed by Eckert and Strube (1999). Most of those false starts were not utterances in themselves. For instance, in (8), the speaker introduces *te* ('you'), but then changes her mind, and produces a different sentence. The entity *you*, however has already been introduced, and therefore it has to be considered as part of the Cf list.

(8)   bueno. <te> /mm/ entonces quedamos así.

      'Good. you mm then (we) agree on that.'

   Following Byron and Stent (1998), "empty utterances", that is, utterances that contain no discourse entities, are attached to their preceding or following utterance, according to context. This applies to empty utterances across turns as well, so that backchannels (Yngve, 1970) are ignored for Centering purposes. (9b) is a backchannel signal, making (9a) and (9c) the adjacent utterances for Centering.

(9)   A: a. Me levanto a las siete

            '(I) get up at seven'

      B: b. Sí.

            'Yes'

      A: c. empiezo las clases de ocho a nueve cuarenta

            '(I) start class from eight till nine forty'

3.3 Linearity and overlapping

A conversation is the combined effort of two or more participants. Reference passes back and forth between speakers, producing a sense of coherent whole for the entire conversation. As a consequence, I considered that Centering transitions applied from one utterance to the next, regardless of whether the two utterances were produced by the same speaker or by different speakers, in line with Byron and Stent's (1998) proposal. This applies when the turns are actually floor-holding (Edelsky, 1981), rather than backchannel signals, as in Example (9) above. Example (10) shows two turns. The centers in B's turn include an entity in A's turn, a reference to B herself.

(10) A: a. qué tal te viene?

            'how is (that) for you?'

            Cf: MEETING (null), B (*te*, 'you')

B: b. no. te contesté recién que /eh/ hoy viernes yo no puedo.

'no. (I) just told you that uh today Friday I can't.'

Cf: B ('I', null), A (*te*, 'you'), FRIDAY

Cb: B

3.4 First and second person pronouns

Spoken language usually contains a high number of first and second person pronouns. Centering was devised explicitly with third person pronouns in mind, and most applications of Centering do not take first and second person pronouns into account. Byron and Stent (1998) found that it was necessary to include them in the Cf list. This is certainly the case in the data, where the antecedent for null first and second person pronouns is to be found in previous utterances. In the following example, *I* and *you* in (11b) are linked to *we* in (11a). Of course, part of that reference is situational, but it can certainly be included in a Centering analysis.

(11) a. Mónica. /eh/ te parece que nos juntemos algún día en la mañana, toda la mañana entera? y trabajemos?

'Monica uh what do you think (we) get together some day in the morning, all morning? and work?'

b. así que querría saber si vos el miércoles diecisiete podés.

'So (I)'d like to know if you can Wednesday the 17th in the morning.'

First and second pronouns, in this data, constitute a large number of the entities for each utterance; in fact, the only entities in many cases. Were they not included in the Cf list, we would find many more instances of transitions with no backward-looking center.

4. What is salient? And how much?

Once the conversations have been segmented into utterances to be considered for the Centering analysis, the next step is to assign a Centering structure to each one of them. This involves (i) building the Cf list (the list of forward-looking centers), (ii) determining the Cb, and (iii) establishing which transition holds between two consecutive utterances. The thorniest of those tasks is the construction of the Cf list. In this section, I discuss the different issues involved in populating the Cf list.

4.1 Entity realization

In Centering, the list of forward-looking centers is a partial ordering of the entities realized in the utterance. Precisely what the definition of 'realized' is, and what criteria we should use for that ordering are the two problems in ranking the Cf list, that is, in deciding which entities are salient in the discourse, and how salient they are in relation to each other. The definition of 'realize' depends, according to Walker, Joshi & Prince (1998: 4), on the semantic theory one chooses. But, in general, "realize describes pronouns, zero pronouns, explicitly realized discourse entities, and those implicitly realized centers that are entities inferable from the discourse situation".

Cornish (2005) argues, in general, that entities in focus are not only those that have been explicitly introduced in the discourse. We need to consider, then, inferable entities. Inferable entities are of particular importance in dialogue because it relies more than monologue on the context outside the text proper. To populate the Cf list, indirect realization of entities was permitted: null subjects; member-set relations (*Mom-Mom and Dad*) and part-whole relations (*branches-trees*). A strict direct realization (where the entities have to be mentioned explicitly in the utterance) resulted in a large number of empty Cbs. What exactly an indirectly realized entity is may, of course, not be obvious. I used the relations identified by Halliday and Hasan (1976) as lexical cohesion (synonymy, hyponymy, superordinate, but not collocation, which does not necessarily involve reference to the same entity). Particularly difficult in this respect were decisions having to do with dates and times, and how those are related to each other. I considered mostly 'include' relations (Hurewitz, 1998), such that, for instance, a date was deemed to be related to the previous utterance's Cf list if it was part of a date range mentioned there. However, when the date was not within the time frame established, it is plausible to think that the hearer had to construct a new model for it. In Example (12), speaker A proposes the week of the fourth, after having discussed the previous week. However, speaker B returns to the previous week, and mentions Friday, October 1st, i.e. a date not in the week of the fourth. This is a new entity, and cannot be related to the immediately preceding utterance. As it happens, this results in a empty Cb, since there are no entities in common between the two utterances.

(12) A: … quieres tratar la semana de cuatro?

   '... do you want to try for the week of the 4th?'

   B: qué te parece el viernes primero de octubre, luego de las once de la mañana?

   'what do you think of Friday October 1st, after 11am?'

Spoken language tends to leave much unsaid. That characteristic poses further problems for an account of the 'realize' constraint in Centering. It has been proposed that bridging inferences (Clark, 1977) can be used to relate entities between utterances. In Example (13a), speaker A mentions *Internet*, which is continued in (13b) and (13c), in two null subjects. In (13d), speaker B does not refer to *Internet* at all, but introduces *computer* in the conversation, with a definite article. Usually, there would be no connection between (13c) and (13d): the Cf list for (13c) includes only INTERNET, and the Cf list for (13d) is: B (THE SPEAKER), COMPUTER. However, *computer* is an inferable (Prince, 1981), a computer being needed to access the Internet, and it can therefore become the Cb of (13d), picking up on *Internet* in (13c).

(13) A: a. estoy conectado con Internet y todo

   '(I)'m connected to Internet and all.'

   B: b. qué tal

   'How's (that)?'

   A: c. es bárbaro

   '(It)'s great.'

   B: d. yo no me pude comprar la máquina todavía, loco

   'I haven't been able to buy the computer yet, man.'

Example (14) shows another instance of an inferable entity. The speaker in (14a) says that he wrote 'a lot' (*muchísimo* is an adverb). In the next utterance, he says that '(they) don't arrive'.

The plural null pronoun can be interpreted as being a reference to the product of his writing, probably letters. The two utterances were considered to have LETTERS in common, which is then the Cb of (14b).

(14) a. Escribí   muchísimo,
        write:1SG.PAST   very.much

    'I wrote a lot,'

    b. lo   que   pasa   es que no   llegan.
       the what happen:3SG.PRES is that not arrive:3PL.PRES

    'what happens is that (they) don't arrive.'

Null, or zero, subjects, are common in Spanish, but always recoverable from the context and the morphology of the verb, and are always added to the Cf list. Ambiguous cases do occur, just as pronouns in English can be ambiguous. Those are disambiguated to the most plausible referent when creating the Cf list. There exist other instances of implicit entities, beyond null subjects. In Example (15), the conversation is clearly about children, those of both interlocutors. However, children are only mentioned once, in the first turn. We have to assume that they are implicit in the rest of the exchange, as are the subjects, so that the sentences read: *Do you have children?* and *We don't have children yet.* The summary in (16) represents the two lists of entities of the exchange, depending on a literal interpretation, or one that allows inferable entities[9]. I decided to use the one on the right, which includes all the entities inferable from the context. It is plausible to assume that those entities are in the focus of attention throughout the exchange.

(15) B: a. ... ¿Y chicos?

            'And children?'

    A: b. Sí. Todavía no

            'Yes. Not yet.'

    B: c. ¿Ah?

            'Huh?'

    A: d. Todavía no

            'Not yet.'

    B: e. ¿Todavía no?

            'Not yet?'

    A: f. ¿Ustedes?

            'You (plural)?'

    B: g. Ah bueno, dos ya

            'Ah well, two already.'

---

[9] There is one further complication in (15), a request for repetition in (15c). In (16) I have excluded that turn, since it does not contain entities, under either view (with or without inferable entities).

(16) **Dialogue**

	Cf list without inferables	Cf list with inferables
B: And children?	children	A, children
A: Yes, not yet	-	A, children
A: Not yet	-	A, children
B: Not yet?	-	A, children
A: And you?	B	B, children
B: Ah well, two already	2 (children)	B, 2 children

4.2 Cf ranking

The ranking of the entities in the Cf list is most often performed by following grammatical relations. Thus, subjects are ranked higher than objects, and these higher than adverbials. In English, this results in the following order (Walker et al., 1998):

(17)   Subject > Object(s) > Other

The ranking is, however, not fixed, and considered to be language-dependent. When a new language is considered, a *Cf template* (Cote, 1998) for that language needs to be developed. Several languages have been studied using Centering, and thus different templates exist. For instance, the template for Japanese includes topic markers (*wa*) and empathy markers on verbs, resulting in the following template (Walker et al., 1994).

(18)   (Grammatical or zero) Topic > Empathy > Subject > Object2 > Object > Others

Di Eugenio (1998) also ranks empathy highest in her template for Italian, following Turan's (1995) for Turkish. Turan and Di Eugenio take the notion of empathy from Japanese, and view it as reflected in psychological verbs (*interest, seem*), perception verbs (*feel, appear)* and certain expressions that refer to point of view (*in her opinion*). There are proposals to incorporate other factors in the Cf template, such as Strube and Hahn's (1999) use of discourse status, whether hearer-old or hearer-new (Prince, 1981), to analyze German. Cote (1998) uses Jackendoff's (1990) Lexical Conceptual Structures. Gordon, Grosz and Gilliom (1993) discovered that both grammatical function and surface order had a role in giving an entity prominence within the Cf. In the next few sections I discuss some of the factors that affect Cf ranking in Spanish.

4.3 Empathy and animacy

Spanish is a pro-drop language; subjects do not need to be realized as pronouns if they are known in context. Additionally, it has direct and indirect object clitics (unstressed pronouns). Corresponding stressed object pronouns are possible for animate entities only. I mainly follow grammatical relations as the basis for ordering the Cf list in Spanish. Therefore, subjects are ranked higher than objects, whether they appear as full pronouns, or as null pronouns.

There are two other criteria that play a role in the Cf ordering in Spanish: empathy and animacy. Following Di Eugenio (1998), I take empathy with the speaker or hearer over strict word order as a ranking criterion. Empathy, as defined by Kuno (1987: 206), "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."

There are no studies, to my knowledge, of how empathy and point of view are expressed in Spanish, in general[10]. The main place where I observe empathy-related effects is in the argument structure of psychological verbs. In those, 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 the like). In (19) the speaker is the highest-ranked entity, because it is the experiencer of a psychological verb (*parece*). In this case, the experiencer is encoded with clitic doubling (Fernández Soriano, 1999; Suñer, 1988): the PP *a mí*, plus the clitic *me*. In Example (20), the clitic *me* refers to the speaker, for whom Thursday is a better date[11].

(19) a   mí   me   parece   también, bueno de hacer una reunión,
     to   me   CL.1SG seem:3SG.PRES too   good   of do:INF a   meeting,

'It also seems good to me to have a meeting,'

Cf: I (*a mí, me*), TO HAVE A MEETING, MEETING

(20) me   viene   mejor el   jueves,
     CL.1SG come:3SG.PRES better the   Thursday

'Thursday is better for me.'

Cf: I (*me*), IT (the meeting, null), THURSDAY

However, the point of view criterion need not apply to the speaker only. In (21), the point of view is that of the interlocutor.

(21) este qué tal   para ti,   del   quince al   diecinueve.
     so   how -   for   you:SG from.the fifteen to.the   nineteen

'So, how is it for you from the fifteenth to the nineteenth?'

Cf: YOU (*para tí*), IT (the meeting, null), FROM THE 15TH TO THE 19TH

A number of verbs in Spanish follow this pattern ("me conviene", "me viene mejor", "se me hace que", *it's good for me, it's better for me, it seems to me*). Thus, for these verbs, the thematic role of experiencer takes precedence over the grammatical function of subject. Empathy also includes verbs with clausal grammatical subjects, but with an animate experiencer, or person from whose point of view the statement is to be interpreted. In (22), there is a displaced clausal subject, 'to meet with you that day'. The subject is included in the Cf list as a single entity. The speaker is the most salient entity, represented in *para mí* 'for me'.

(22) así   que   para mí sería   imposible juntar-me   con vos   /eh/ ese día
     so   that for   me be:PRES.COND impossible join:INF-CL.1SG with you:SG uh   that day

'So it would be impossible for me to meet with you that day.'

Cf: I (*para mí*), TO MEET WITH YOU, THAT DAY

---

[10] Although Wanner (1994) and Heap (1998) discuss how empathy affects the ordering of clitics in Spanish.

[11] Abbreviations used in the examples: 1/2/3 – first/second/third person; CL – clitic; NOM – nominative; ACC – accusative; DAT – dative; SG – singular; PL – plural; FEM – feminine; MASC – masculine; POSS – possessive; PRES – present; PRET – preterite; INF – infinitive; GER – gerund; SUBJ – subjunctive; COND – conditional.

Not all experiencers, however, seem to be good candidates for higher placement. In a sentence like *Juan asusta a María,* 'John frightens Mary', the subject Juan seems to me to be more prominent than María, although María is an experiencer. It is possible that experiencers are ranked higher only when they are first and second person, which also happen to be higher in most hierarchies of animacy[12].

Animacy is a relevant feature in the ordering of clitics and reflexive pronouns that refer to participants in the discourse. Animacy is considered relevant in general for salience and topicality (Givón, 1983). Stevenson et al. (1994) found that animacy has a role in deciding which entity will be in focus, and it was also found to have an effect in pronominalization (GNOME, 2000)[13]. Clitics and reflexive pronouns, in addition to conveying empathy (see above), are also placed before the verb, linearly before (clitic) direct objects (whether empathy is involved or not)[14]. It is usually the case that indirect objects are animate, whereas direct objects may not be. In summary, three reasons speak for ordering the objects as indirect before direct: (i) indirect objects can convey empathy; (ii) indirect object clitics are always placed before direct object clitics; (iii) indirect objects tend to be animate. Wanner (1994) argues that clitic sequences in Spanish obey constraints of empathy and animacy. An illustration is to be found in (23), where 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. Notice that the null subject is arbitrary (see below), and thus ranked last.

(23) a. Mi hermana solicitó un programa de arqueología y antropología en Grecia.

'My sister applied to a program in archeology and anthropology in Greece.'

b. ¡Y que se lo dan!
and that CL.3SG.DAT CL.3SG.MASC.ACC give:3PL.PRES

'And they give (gave) it to her!'

Cf: SISTER (*se,* 'to her'), PROGRAM (*lo,* 'it'), THEY (null)

4.4 Cf proposal for Spanish

Subjects take precedence in the Cf list in most other cases (i.e., when they are not clausal, and when there are no experiencers). Accordingly, the elements of the Cf list follow the order in (24)[15]. This ranking applies first to main (matrix) clauses, and then to subordinate clauses, when the two are within the same Centering unit (usually, because the subordinate clause is non-finite; see Section 3.1 on segmentation).

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

At the end of the ranking are null arbitrary subjects (Jaeggli, 1986), as in (23) above, and subjects in impersonal constructions with *se,* as in Example (25). The word *se* in this example

---

[12] Thanks to Jeanette Gundel and Nancy Hedberg for bringing up this point and suggesting the example.

[13] Zaenen et al. (2004) discuss previous literature on the importance of animacy in a number of areas, including the choice between Saxon genitive and the *of*-genitive, which may affect ranking in Centering.

[14] See Heap (1998) for an Optimality Theory account of how empathy is also involved in non-standard rearrangements of clitics.

[15] This Cf template is slightly different from previous proposals (Taboada, 2002a, 2002b).

indicates a non-specific subject in an impersonal middle voice construction (Mendikoetxea, 1999), meaning "one can hear that you are well".

(25) Ya   se   te   oye   muy bien.
already *se* CL.2SG.ACC hear:3SG.PRES very well

'You already sound very well.'

Cf: YOU (*te*), ONE (*se*)

Also included as impersonal pronouns are instances of the second person singular, which can be used impersonally (Butt and Benjamin, 2000). It is interesting to note that this second person form is often used as an indirect form of reference to the speaker. In Example (26), the speaker is implying that he has to take one exam every year. The *tú* form might indicate simply that that's the norm, and he is no exception. If we were to consider that the second person form has some reference to the speaker, its ranking in the Cf list would have to change to: I (SPEAKER), EXAMS, EVERY YEAR, since the subject is the second person singular. The sentence, however, seems to be more about the exams than about who has to submit them.

(26) a. Son, son los tutoriales.

'(They) are the exams.'

    b. Tienes   que presentar uno   cada año.
        have:2SG.PRES that submit:INF one   every year

'(You) have to submit one every year.'

Cf: EXAMS (*uno*), EVERY YEAR, ONE/YOU (null subject)

4.5 Noun phrases with more than one entity

A few other issues need to be addressed in the Cf ranking. The first is related to noun phrases that contain more than one referent or entity, whether possessives (*my brother, my letter*), nouns with a prepositional phrase (*the census of the city*), or conjoined NPs (*Juan and María*). For possessives I follow Di Eugenio (1998): the possessor is ranked before the possessed, if the possessed is inanimate, and the possessor after the possessed, if the possessed is animate[16]. In (27), the ranking of *mi examen* ('my exam') is SPEAKER > EXAM. However, in (28), the ranking of *mi mamá* ('my Mom') is MOM > SPEAKER.

(27) Una maestra este, me   tuvo   que venir   a
    a   teacher eh   CL.1SG have:3SG.PRET that come:INF to

        hacer   mi último examen aquí.
        make:INF my last   exam   here

'A teacher uh, had to come and give me my last exam here.'

---

[16] Gordon et al. (1999) suggest that the head of the NP (i.e., the possessed) is always the most salient. However, their experiments were based on NPs with animate possessor and possessed. The experiments were designed to test (and debunk) a linearity hypothesis (Gernsbacher and Hargreaves, 1988; Walker and Prince, 1996), but they were all conducted in English. Further crosslinguistic experiments would be desirable: in Spanish, and in other languages, possessives with two full NPs (e.g., *Mary's letters*) have a different word order (*las cartas de María*). Tetreault (2001) shows that an anaphora resolution algorithm performs better using Gordon and colleagues' ranking—though the corpus used was English as well.

Cf: TEACHER, I (*mi*), EXAM, HERE

(28) mi   mamá   posiblemente llegue   la   otra semana
     my   Mom   possibly   arrive:3SG.PRES.SUBJ the   other week

    'My Mom will probably arrive next week.'

    Cf: MOTHER, I (*mi*), NEXT WEEK

The same principle applies to noun phrases with a PP modifier usually headed by 'of' (*de* in Spanish). In most of those constructions, the meaning is that of a genitive (*las cartas de Marta* = Marta's letters). The approach taken here is different from Walker and Prince's (1996) Complex NP Assumption, which ranks NPs with a possessive determiner in linear order, left-to-right. Since I are considering animacy as a relevant feature, I preferred to follow Di Eugenio's ranking for possessives, and to expand it to other NPs that include more than one entity. Thus, in Example (29), *una de Marta* refers to one (letter) from Marta. Since *Marta* is animate, it is ranked higher than *letter*.

(29) Y   una   de Marta.

    and one   of Marta

    'And one (letter) from Marta.'

Conjoined NPs activate as most salient entity the group denoted by the conjoint. Thus, in *John and Mary,* the most salient entity is the group JOHN AND MARY. The individual entities, JOHN and MARY, are less salient than the group (Gordon et al., 1999). In that same paper, Gordon and colleagues suggest that the individual entities are equally salient. The mention of either JOHN or MARY results in the same processing time in a psycholinguistic experiment. It could be argued that this result would lead to multiple entities in the same position within the Cf list, as in (30), where the separate entities JOHN and MARY occupy the same place in the Cf list. However, I feel that allowing multiple entities in the same position would make ranking too complex, and would also complicate future attempts at implementing these methods in an anaphora resolution system[17], and prefer to use linear order to sort the two entities (31).

(30) John and Mary went to the store.

    Cf: John and Mary, John, Mary, store

(31) Cf: John and Mary, John, Mary, store

4.6 Wh-pronouns

Wh-pronouns, *qué* ('what'), *quién* ('who'), *cuándo* ('when'), are included in the list of forward-looking centers, and are ranked according to the syntactic role they have in the clause. Although wh-pronouns do not have a specific referent, they do serve as antecedents for other referring expressions. According to Halliday (Halliday, 1967; Halliday and Matthiessen, 2004), wh-words can be Themes in a clause, and I believe that they can establish cohesive ties throughout a text[18].

---

[17] Poesio et al. (2004) discuss the need for a second criterion when two entities may be ranked in the same place. They use linearity.

[18] Pesetsky (1987; 2000) proposes that some wh-words are D(iscourse)-linked, that is, they ask a question whose answer is drawn from a salient set. However, he says that only *which* questions are D-linked. I think that all wh-

In (32b), *qué* 'what' is included in the Cf list, and used as an antecedent for *ecología* in (32c), thus becoming the Cb of that utterance.

(32) B: a. se va a la Universidad de Gales, del Sur, donde estudió Sarucán, también.

'She is going to the University of South Wales, where Sarucán studied as well.'

  A: b. A hacer qué.

'To do what?'

  B: c. Este. A hacer ecología.

'Eh, to do environmental science.'

4.7 Reference through more than one expression

An utterance may contain reference to the same entity through more than one referring expression. For instance, the utterance in (33) contains reference to the subject both through the null subject pronoun and through a clitic (*nos*). In Centering we are usually concerned with the entities mentioned in the utterance, not so much with the referring expression(s) used to evoke them. However, since my concern in this paper is the link between Centering transitions and referring expressions, this was an important issue. The ranking of such entities is straightforward: the most salient grammatical function (or other criterion that may apply) is used to list the entity in the Cf list. The problem is which form should be used to categorize the form of the Cb in that utterance (see Table 4 below). I have, for the time being, categorized such examples under the most marked form of reference. In Example (33), the referring expression used to denote the entity "first person plural" is listed as a clitic, not as a null pronoun (clitics are considered more marked than null pronouns). It could be argued that the least marked form should be used to classify the Cb, but that would not show the fact that the Cb is, in a way, reinforced by another referring expression, by being referred to twice in the same utterance.

(33) nos vamos   con mi madre
  CL.1PL go:1PL.PRES with my mother

 '(We) are going with my mother.'

 The verb *be* (*ser* and *estar* in Spanish) functions as a linking verb, so subjects and predicates (nominal and adjectival) of the verb *to be* are coreferential and only need to be listed once in the Cf list. In (34), there are two references to the person the speaker is talking about, his teacher. The first reference is through the null subject, and the second through the predicate noun, *amiga.* The case is similar to the one above, where two referring expressions are used to refer to the same entity. As above, I classified the most marked one (NP in this case).

(34) porque aparte es   mi amiga,
  because besides be:3SG.PRES my friend:FEM.SG

 'because (she)'s also my friend,'

 It is possible to have only a predicate (elliptical subject and predicator) in an utterance. In these cases, since the predicate is coreferential with the elliptical subject of the elliptical predicator, I include the subject in the list of forward-looking centers. In Example (35), the

---

words establish a link between the question and its answer.

speaker refers to himself with 'covered'. Although there is no predicator in the sentence, reference to the speaker is included as if a null subject were present.

(35) Lleno   de granitos, no,   este
    full:MASC.SG   of zits   no   eh

    '(I'm) covered in zits.'

    In most cases, the subject and the nominal predicate have exactly the same reference. In some cases, the reference may be slightly different: *The dinner choice is pasta*[19]. Miltsakaki and Kukich (2004) label these predicates as *specificational* (and predicates such as the one in Example (34) as *predicational*). They rank specificational predicates higher than their corresponding subjects. I did not make such distinction, and treated all linking verb predicates in the same manner, as described above: the first (subject) reference determines the location in the Cf list; the predicate determines the type of referring expression used to refer to the Cb, if the entity in question is the Cb of the utterance.

4.8 Right and left-dislocation

The ordering of the Cf list is affected by other factors, among them right and left-dislocation. I have not, for the moment, dealt with those, but a closer look at the data suggests that the ranking will be affected by dislocated elements. In Example (36), two different rankings are possible. The first one (37) ranks *modem* according to its grammatical function, object. The alternative (38) is to rank it higher than the *pro* subject *we*, because it is left-dislocated. The usual ranking produces a RETAIN transition from (36a) to (36b), and a SMOOTH SHIFT from (36b) to (36c). The alternative ranking, with *modem* higher, results in a CONTINUE followed by a RETAIN[20].

(36) A: a. ¿módem?

        'Modem?'

    B: b. módem, los tenemos

        'Modems, (we) have them'

      c. pero no los instalamos todavía

        'but (we) haven't installed them yet.'

(37) Grammatical ranking

    a. Cf: MODEM

      Cb: 0

    b. Cf: WE, MODEMS

      Cb: MODEMS – Transition: RETAIN

    c. Cf: WE, MODEMS

---

[19] Thanks to Laurie Fais for this point and for the example.

[20] Transition preference for individual utterances is perhaps not enough of a reason to consider the alternative. Rule 2 is mostly about preference for sequences of certain transitions. Another complicating factor is that left-dislocation may not signal salience: Givón's (1983) topic accessibility scale ranks left-dislocated NPs as less accessible than neutral-ordered NPs. It is not clear whether less accessible in Givón's scale means more salient in Centering terms.

Cb: WE – Transition: SMOOTH SHIFT

(38) Alternative ranking

    b. Cf: MODEMS, WE

        Cb: MODEMS – Transition: CONTINUE

    c. Cf: WE, MODEMS

        Cb: MODEMS – Transition: RETAIN

4.9 Unresolved issues

There are a number of unresolved issues in the ranking of the Cf list. The first one is the use of prosody in addition to the other factors that affect the ranking. A number of researchers have pointed out that prosody and stress affect the order of elements in the list when dealing with spoken language (Brennan, 1995; Cornish, 1999). This remains a task to be addressed in future research. Another difficulty within Centering is the treatment of pronouns that refer to discourse segments, or to abstract entities (Asher, 1993; Byron, 2002). I have excluded them from analysis for the time being.

5. Which referring expression?

Anaphora resolution, and the form of the anaphoric term itself have long been linked to the relative prominence of entities in the discourse (Gundel et al., 1993; Prince, 1981; Sidner, 1983). Rule 1 of Centering Theory establishes that the Cb of an utterance must be a pronoun, if other pronouns are present. That is, the Cb will be realized by the most reduced form (a pronoun) if other pronouns are present. Centering does not suggest any other rules for what will happen in other situations, i.e., when there are no pronouns at all. However, researchers have proposed a relation between the transition type, i.e., the progression of local discourse topics, and either the form of referring expressions used to realize the subject (Di Eugenio, 1998), or the Cb of an utterance (Taboada, 2002a). The main purpose of this paper is to determine what relationship there is between Centering transitions and referring expressions. For that purpose, I carried out a corpus analysis of two types of spoken language corpora in Spanish. The corpora are the ISL corpus and the CallHome corpus.

The ISL corpus is a large collection (a total of about 500 conversations) of task-oriented conversations recorded in a lab, with externally-controlled turns. The participants, who were native speakers of Spanish[21], had to press the 'Enter' key on a keyboard to yield the turn, which makes the conversations similar to one-way radio, although the speakers are present in the same room. The task was to arrange for a two-hour meeting within a time period that ranged from two to four weeks. The speakers had conflicting agendas, and usually proposed a number of dates before an agreement was reached. Nine conversations from this corpus were analyzed, three each of dyads of female-female, male-male and female-male speakers. The nine conversations amounted to 262 utterances, as defined in Section 3.1, and a total of 2,798 words.

---

[21] The speakers came from all corners of the Spanish-speaking world. For more details on the corpus, see Taboada (2004).

The CallHome corpus is a collection of telephone conversations lasting up to 30 minutes between native speakers of Spanish. One party was given a free long-distance call, free choice of who to call, and no restriction on topics. Most participants called relatives or friends[22]. For this study, five conversations were used, a total of 1,198 utterances and 8,694 words.

The conversations were first segmented according to the guidelines outlined in Section 3.1. Then each utterance was coded according to Centering principles, including Cf list and type of transition. Table 2 shows the number of non-zero transitions for both corpora.

	Continue	Retain	Smooth shift	Rough shift	Total
ISL	121 65%	27 16%	22 15%	11 4%	181
CallHome	515 67%	129 15%	116 12%	30 6%	790

Table 2. Centering transitions in two corpora.

The results are as predicted in Centering: CONTINUE transitions are preferred (overwhelmingly) over other types of transitions, and RETAINS are preferred over SHIFTS. ROUGH SHIFTS are relatively rare. It is interesting to see that the two corpora have similar percentages of all types. Although the corpora are both spontaneous spoken conversations, they are somewhat different, in that the ISL conversations are task-oriented, whereas the CallHome recordings are casual. Those differences do not seem to affect the distribution of Centering transitions.

The numbers shown in Table 2 are for transitions that had a backward-looking center. A large number of transitions had an empty Cb, and were not included in the analysis. The numbers are presented in Table 3.

	Utterances	Cb=0	%
ISL	262	81	30.92%
CallHome	1198	408	34.06%

Table 3. Utterances with empty backward-looking centers.

There exist a number of reasons for the high occurrence of utterances with an empty backward-looking center. Some of those utterances do introduce completely new entities in the discourse, thus beginning a new discourse segment: Centering operates at the local discourse level; transitions between discourse segments are part of the global structure, and strictly not part of a Centering analysis[23]. In a number of cases, however, the entities were inferable from the

---

[22] Participants were also speakers of different dialects. Details about the transcriptions are available at:

http://www.ldc.upenn.edu/Catalog/docs/LDC96T17/ch_span.txt

[23] Identifying discourse segments is not a trivial matter. My observations here about when discourse segments start are impressionistic; rigorous analysis and annotation needs to be done to integrate Centering into the global structure of the discourse.

context, but the inference seemed a bit far-fetched, and I decided not to establish it. That is the case in (39b), where the speaker refers to the days she has mentioned in utterance (39a). The utterance could read "check if you can meet on Tuesday the 16th after 12 noon", but instead it is "check if you can". This is not just a question of a null object, but a null VP. I decided to not include the date in the Cf list for (39b).

(39) a. así que recién podría el martes /eh/ dieciséis después de las doce del mediodía.

     'So I could on Tuesday, uh, the 16th after 12 noon.'

  b. fijate si vos podés.

     'Check if you can.'

As I pointed out in Section 4.1, the issue here is what kind of inferables can be included in the Cf list of an utterance. Hurewitz proposes to include entities that are in a functional dependency with previously mentioned entities or that are subsets of other entities (Hurewitz, 1998), and also discourse deictic pronouns, i.e., pronouns that refer to a part of the discourse, such as events or clauses (Webber, 1981). In Hurewitz's account, utterances joined by one of those relations constitute a new type of transition, a PARTIAL SHIFT. Fais (2004) links entities in the discourse to other previously mentioned entities using cohesive relations (Halliday and Hasan, 1976).

In some other cases, empty Cbs resulted from problems with the segmentation (Poesio et al., 2000), or from the strict adjacency constraint in Centering: only entities in the previous utterance can become the Cb of the current one. Some empty Cbs were as predicted by Centering, that is, they initiated a new discourse segment; for instance, a new topic is being discussed, or a new date is being proposed[24], and therefore contained no link to the previous utterance. The new discourse segments are often a completely new 'push' onto the focus stack (Grosz and Sidner, 1986), but they can also be insertion or side sequences (Jefferson, 1972) or corrections (Schegloff et al., 1977).

5.1. Referring expressions and transitions

The Cb of each utterance was coded according to whether it was one of the several possible referring expressions, and those types of expressions were related to the transition types. The referring expressions are illustrated in (40) to (46). The referring expression in question is in bold.

Zero pronoun

(40) a. Conozco, en serio, un doctor que hizo su doctorado en Japón,

     'Seriously, I know a doctor who did his Ph.D. in Japan,'

  b. acabó   y   [pause]
    finish:3SG.PRET and

     '(He) finished and'

  c. no **[-]**   encontró   chamba,

---

[24] See Taboada (2000; 2004, ch. 6) for a discussion of discourse segments in the ISL conversations. A new discourse segment was always initiated when a new date is being proposed, as evidenced by a break in the chain of cohesive links in the conversation.

not **null** find:3SG.PRET employment

'didn't find a job,'

Clitic

(41) a. Llega a Atenas

'(She) will arrive in Athens'

    b. y va a estar ahí tres semanas

'and (she) is going to be there for three weeks'

    c. y   luego **la**   andan   paseando de isla en isla
and then **CL.3SG.FEM.ACC** go:3PL.PRES walk:GER from island to island

'and then (they) are going to take her around from one island to the next.'

Pronoun

(42) a. No, no. Si de hambre no me muero.

'No, no. (I)'m not going to starve.' (lit., 'die of hunger')

    b. Pero **yo** quiero   ser   astrofísica
but   **I** want:1SG.PRES be:INF astrophysicist

'But I want to be an astrophysicist.'

Demonstrative pronoun

(43) B: a. Aquí le llaman tutorial.

'Here they call it tutorial.'

   A: b. Sí, pues ha   de ser   **eso**.
yes then have:3SG.PRES of be:INF **that**

'Yes, then it must be that.'

Full noun phrase

(44) B: a. ¿Y tu hermana?

'And your sister?'

   A: b. **Mi**   **hermana** está   bie-
My   **sister**   be:3SG.PRES we-

'My sister is well.'

Other

(45) Wh- pronoun

   A: a. Ay, pero no muchos días más.

'Ah, but not many more days.'

   B. b. **Cuánto** más.

'**How much** more?'

(46) Adverbial (NP or PP)[25]

> A: a. no. el lunes en la mañana \<no\> no puedo.
>
> 'No. Monday morning (I) can't.'
>
> b. tal vez el lunes en la tarde, después de las doce?
>
> 'Maybe Monday in the afternoon, after twelve?'
>
> B: c. bueno **el lunes** tengo una reunión de \<d\> uno a cuatro
> well **the Monday** have:1SG.PRES a meeting from tw- one to four
>
> 'Well, on Monday (I) have a meeting from tw- one to four.'

	Continue		**Retain**		**Smooth shift**		**Rough shift**	
Zero pronoun	350	55.0%	44	28.2%	74	53.6%	9	21.9%
Clitic	114	17.9%	48	30.8%	24	17.4%	14	34.1%
Pronoun	53	8.3%	8	5.1%	9	6.5%	4	9.8%
Demonstr. pr.	15	2.4%	4	2.6%	4	2.9%	4	9.8%
Full NP	86	13.5%	26	16.7%	22	15.9%	9	21.9%
Other	18	2.8%	26	16.7%	5	3.6%	1	2.4%
n	636		156		138		41	

Table 4. Referring expressions for the Cb of each utterance, according to transition.

Table 4 shows that, overall, the Cb tends to be expressed through a zero pronoun. This is the least marked form available in Spanish. For that reason, it is to be expected that the Cb will be coded as a zero pronoun when the transition is a CONTINUE. Such is the case: out of the 636 continue transitions (for both corpora together), 55% had a zero pronoun as Cb. When we move onto RETAIN, where the Cb is continued from the previous utterance, but will likely not be continued further, the percentage of zero pronouns decreases. However, it grows again in the SMOOTH SHIFTS, to almost the same percentage as for CONTINUE (53.6%).

Di Eugenio (1990; 1998) found that in Italian[26], speakers typically encode center continuation with zero subjects, and center retention and shift with stressed pronouns. She also found that instances of RETAIN and SHIFT with null pronoun subjects are possible if the utterance that constitutes the change contains syntactic features that force the zero subject to refer to an entity other than the Cb of the previous utterance. Indeed, I found many cases of null pronouns in subject position that made the referent clear, when it was other than the Cb_{i-1}. In Example (47c), the number agreement on the verb links the null subject to the object in the previous utterance ('mountains'), not its subject and Cb, the Yosemite National Park that the speakers have been discussing.

(47) a. Sí. Sí, es un parque nacional

---

[25] Adverbials that are added to the Cf list are mostly those that denote times and places.

[26] Di Eugenio analyzed excerpts from two novels, newspaper articles, short stories, and a bulletin board post. There is a difference between Di Eugenio's analysis and mine: she studied the realization of the subject; I examine the Cb.

'Yes, yes, (it)'s a national park'

b. y es, tiene así montañas,

'and (it)'s got like mountains,'

c. no, no son muy grandes,

'(they) are not, not very big,'

Di Eugenio also found that speakers encode center retention or shift with a stressed subject pronoun (presumably in the cases when syntactic factors do not exclude reference resolution to the previous Cb). If we look at Table 4, we can see that pronouns are not used very often, across all four transition types. They actually occur less often in RETAIN and SMOOTH SHIFT transitions than in CONTINUE, and only increase within ROUGH SHIFT, to 9.8%, which are only four instances, given the low number of ROUGH SHIFTS.

More numerous are full noun phrases (definite noun phrases or proper nouns), and for those we can see a steady increase from CONTINUE to ROUGH SHIFT. It is possible that center change is expressed more often in (spoken) Spanish via a full noun phrase. For instance, in (48), the conversation has been about B's activities, and she is then the Cb in (48a). When B takes her turn, she shifts and talks about Cristina, previously introduced. She could have used a stressed personal pronoun (*ella*), especially given that there is no competing referent, but instead chose to repeat the proper name.

(48) A: a. Mary, tú fuiste por tu vestido rojo donde Cristina.

'Mary, did you go get your red dress from Cristina's?'

B: b. Mmm. Ay, sí, pero Cristina está en Bogotá

'Mmm. Oh, yes, but Cristina is in Bogotá.'

Clitics are, after null pronouns, the preferred form of realization across transition types. They are used in CONTINUE to refer to the speaker quite often, with psychological verbs (49), or other verbs, as indirect objects (50).

(49) **me**   parece   lo   mejor dejar-lo
    CL.1SG seem:3SG.PRES   the   best leave:INF-CL.3SG.MASC.ACC

      para la   otra   semana,
      for   the other   week

   '(It) seems better to me to leave it for next week,'

(50) para que **me**   lo   arregle   ahora para diciembre.
    so   that **CL.1SG** CL.3SG.MASC.ACC   fix:3SG.PRES.SUBJ   now   for   December

   'So that (she) can fix it for me now for December.'

Clitics do not always refer to the speaker. They can refer to the interlocutor (51) or to a third party, as in (52), with a pronominal verb, *se vino* ('came')[27].

---

[27] The word *se* in this example is a clitic, co-referential with the subject, and different from the *se* in Example (25). This *se* is in a paradigm with other clitics: *me* for first person singular subject; *te* for second person singular subject, etc. These constructions are referred to as pseudo-reflexive or middle voice constructions (Mendikoetxea, 1999). They appear to be reflexive, but are used with intransitive verbs, some of which have both an intransitive and a pseudo-reflexive use (hence the term 'pronominal verbs' when used pseudo-reflexively). See also Sharp (2005) for a

(51) correcto Mónica, **te**   decía   el   viernes
     correct   Mónica   **CL.2SG.ACC** tell:1SG.PRET the Friday

            por aquello de la   muerte de Gaitán,
            for that   of the death of Gaitán

     'Right Mónica, (I) was telling you Friday because of Gaitán's death,'

(52) Sí, **se**   vino   para   acá   estar   conmigo.
     yes **CL.3SG** come:3SG.PRET towards here be:INF   with.me

     Yes, (she) came to be here with me.'

Demonstrative pronouns are not very frequent in general: there are only four instances each for RETAIN and both SHIFTS. They are slightly more common in CONTINUE transitions, but still only account for 2.4% of the Cbs in those. They are used to refer to both things (53) and people (54). In some cases, they also refer to abstract entities, as in (55). Gundel et al. (1993) found that demonstrative pronouns are rarely used for referents that are familiar or in focus, which would be the case with most Cbs in this study, and Ariel (1988) also found a very low level of demonstratives in her corpus analysis.

(53) B: a. lo que sí son buenos, y no los sé usar, son los los enlaces para estar así platicando

         'What's good, and (I) don't know how to use them, are the the links to be like chatting.'

   A: b. Ahá

         'Uh-huh.'

   B: c. **esos** son buenos

         'Those are good.'

(54) B: a. No, eh, ay, mami, el viernes se viene Alicia, la que tú tenías, para acá.

         'No, uh, uh, mami, on Friday comes Alicia, the one you used to have, here.'

   B: b. Vamos a ver qué tal me resulta.

         '(We)'ll see how she turns out.'

   A: c. **Esa** es una fiera.

         'That (one) is amazing.' (lit. 'She's an animal.')

(55) A: a. Aquí aprovechando las llamaditas estas que nos dan, ah

         'Here, taking advantage of these calls (they) give us, uh'

   B: b. Ay, sí, claro, pero eso cómo es, Chipi,

         'Ah, yes, right, but how's that, Chipi,'

      c. **eso** cómo funciona.

         'How does that work?'

Finally, the category "Other" includes a number of other realizations: wh-pronouns, adverbial NPs, and possessive determiners and pronouns. Some of these appear frequently in RETAIN transitions, in preparation for a change of topic. In (56), for instance, speaker A is expressing

---

unified account of all instances of *se* in Spanish.

despair, and ends his turn with a rhetorical question that includes reference to himself in a null pronoun. Speaker B continues talking about speaker A, but uses a possessive determiner (*tu* 'your'), trying to steer the conversation towards exactly what is the problem (*resentimiento* 'resentment').

(56) A: … qué voy a hacer?

       'What am (I) going to do?'

   B: ya, hay mucho resentimiento en tu voz, no?

       'I see, there's a lot of resentment in your voice, isn't there?'

In summary, a CONTINUE transition generally realizes the Cb as a zero pronoun, followed by a clitic. RETAIN transitions are also realized through zero or a clitic, although other possibilities exist. These realizations are as expected, and reflect the types of situations that the different transitions were meant to encode. The next section deals with realizations that appear to be contrary to expectation.

5.2. Realization against expectation

The descriptions in the previous section are all of the type 'x transition *tends to* encode the Cb in y form'. We have seen there are some clear tendencies. My concern here is the realizations that do not follow those tendencies.

The most clearly stated tendency, in this paper and in the literature, is that the Cb of a CONTINUE transition is realized via a reduced expression: zero pronoun, clitic, unaccented pronoun, etc. Other realizations are said to make processing more difficult. For instance, Gordon et al. (1993: 341) establish that there exists a 'repeated name penalty', where repeating a name that continues to be the Cb in the discourse deprives the reader[28] of an important cue that the current utterance is coherent with the previous one. And yet, 13.5% of the Cbs in CONTINUE transitions are realized as full noun phrases, many of them proper names[29].

The explanations for repeated noun phrases all have to do with spoken language phenomena. For instance, in the CallHome corpus, speakers frequently ask about other friends or relatives. These exchanges typically involve one speaker mentioning the name of the person, and the other repeating the name, as in (57). (Proper names are included in the full NP category).

(57) A: Qué han sabido de Eddie.

       'What have you heard from Eddie?'

   B: De Eddie nada,

       'From Eddie, nothing,'

This is quite frequent when the turn changes, but it also happens within a speaker's turn. In (58), the speaker repeats *Mónica*, although the referent should be clear, and the clitic *la* would have sufficed.

---

[28] Gordon et al.'s (1993) experiments were written. My explanations for the lack of 'repeated name penalty' are all related to the fact that the data analyzed here is spoken.

[29] Di Eugenio (1998) found a few instances of strong pronouns in subject position with CONTINUE transitions. She relates it to the transition type preceding the CONTINUE, a possibility I have not yet explored in my data.

(58) a. y Mónica sin embargo ha crecido un montón.

      'And Mónica, however, has grown a lot.'

  b. Tu papá se   asombra   de ver-la   a Mónica,
     yourDad CL.3SG   surprise:3SG.PRES of see:INF-CL.3SG.FEM.ACC to Mónica

      'Your Dad is surprised to see her, Mónica,'

Brennan (1995) found that referents introduced in object position were then re-introduced in subject position with a full noun phrase. Only after that were they referred with a pronoun. Brennan believes that the referent needs to be in subject position so that it can become a backward-looking center, and thus candidate for pronominalization. This is the case in some of the examples, as in Example (57), where the repeated NP/proper name becomes the backward-looking center of the utterance. I also found in the corpus instances of entities in subject position, but left-dislocated (*Y Juan, ¿cómo está?* 'And Juan, how's he?'). The proper name is repeated in subject position before it is pronominalized. It is possible that a neutral subject position is necessary before pronominalization takes place.

In general, 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 noun.

In the ISL corpus, repeated referents across turns are either the participants or the dates being discussed. In (59), speaker B refers to herself with a full pronoun at the beginning of her turn. Amaral and Schwenter (2005) discuss cases like (59), and propose that the pronoun is obligatory, because it establishes a contrast[30].

(59) A: puedes reunirte conmigo en mayo?

    'can you meet with me in May?

  B: a ver **yo** estoy de viaje del   treinta y   uno hasta...
    to see **I** am   of travel from.the thirty   and one until…

    'let's see, I am away from the 31st until...'

In (60), speaker B uses a full NP, *el jueves* to refer to the date being discussed, present in the immediately preceding utterance as a null pronoun. Note that in this case, contrast does not play a role.

(60) A: a. creo que el jueves veintisiete, que lo tengo totalmente libre podría ser.

      'I think Thursday the 27th, which (I) have completely free, it could be.'

    b. qué te parece?

      'What do you think (of that date)?'

---

[30] Dimitriadis (1996) proposes that a pronoun is chosen when the antecedent is not the Cp of the previous sentence (i.e., it is not the most salient entity in the previous sentence). It is possible that that is the case in many situations, but not in Example (59), where *tú* ('you'), the null pronoun from the first utterance is realized as a strong pronoun (*yo*) in the second utterance, of course with the change in person due to the change of speaker. Contrast and the change of turn seem to be the decisive factors here.

B: c. bueno. el jueves realmente es un día ocupado para mí.

    'Well, Thursday is actually a busy day for me.'

The presence or absence of the personal pronoun subject in Spanish has received a great deal of attention (e.g., Alonso-Ovalle et al., 2002; Cameron, 1992; Davidson, 1996; Enríquez, 1984). 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. This seems to be the case in the ISL corpus, where the speaker's agenda is contrasted with the interlocutor's. Davidson (1996) finds that the personal pronoun is used for emphasis and to negotiate conversational turns (to claim the floor for an extended period of time). He also found that the first person pronoun was used more frequently than second or third person pronouns. Those three factors might account for the presence of *yo* in examples such as (59).

Conversely, using a zero pronoun when something else is expected could result in more difficulty in processing. In Example (61), speaker A has been talking about visiting his sister in Greece. The Cb at the end of A's turn is *sister*. B then replies with a question, 'isn't (that) very expensive?'. There is no repeated entity across the turns, but B uses a zero for the third person singular subject of (61e). One possible referent is the idea of sightseeing, which A used at the end of his turn. It is possible that B realizes this possible mistaken interpretation, and reformulates in (61f), to specify that she is referring to the cost of the flight, not of doing tourism[31].

(61) A: a. Sí. Sí pues, es que se va a ir a Grecia

    'Yes. Yes, so she's going to Greece'

    b. y luego se queda las tres últimas semanas

    'and then she's staying the last three weeks'

    c. tres, tres semanas más, se queda

    'three, three more weeks, she's staying'

    d. y   no   más   se   pasea
        and not more CL.3SG walk:3SG.PRES

    'and she's just going to do tourism.' (lit. 'she's just going to walk around')

B: e. Pues, ¿no te   sale   carísimo?
        but   not CL.2SG come.out:3SG.PRES   very.expensive

    'Well, isn't (that) very expensive for you?'

    f. o sea el avión, yo digo.

    'I mean the plane, I mean.'

---

[31] Geluykens (1994) attributes this type of repair to a conflict between principles of Clarity and Economy, derived from Grice's (1975) maxims.

6. Conclusions

I have presented an application of Centering theory to two corpora of spoken Spanish. The study contributes to an understanding of the relationship between Centering transitions and choice of referring expression. The analysis shows that, when the topic stays constant, i.e., when a CONTINUE transition is present, the most common realization of the backward-looking center is in a null pronoun. Null pronouns are also used in the other three transition types, likely because they are clearly identifiable from context, through person or number marking.

Full noun phrases and pronouns are used quite often to encode the backward-looking center. This is contrary to the expectation that the topic of the utterance is encoded with the minimum amount of information. According to Gordon and colleagues (Gordon et al., 1993), there is a 'repeated name penalty' when using a more informative referring expression than necessary. It was found that speakers tend to repeat pronouns referring to themselves, and proper names referring to third persons. This occurs most often when there is a change of turn, but also within the turn. In spontaneous conversation, Downing (1996) found that proper names are often used, even when pronouns would ensure correct identification of referent, to establish the referent in the discourse as common to both speakers.

It is possible, then, that we may be forced to revise Centering predictions as to center realization, to take into account spoken language phenomena. Another source of evidence in support of this view is that a number of empty backward-looking centers were attributed to the presence of side or insertion sequences, which are characteristic of conversation. This ties in with the relationship of Centering and the global structure of the discourse. Centering was designed as a model for the local focus of attention. It is not clear how Centering can relate the two levels. For instance, in (62), there is a global story about how speaker B's boss was quite proud of his work in a particular situation, because speaker B and his boss had issued 2,700 notices for back taxes on cars. Speaker B then starts a small story about how his boss came to know that he had done well in comparison to others. The story covers utterances (62c) to (62j). In utterance (62k), speaker B refers again to his boss, which is part of the global focus of discourse, as part of the 'we' in (62a). However, in (62k) the subject *el tipo* 'the guy' cannot be linked to the immediately preceding utterance, and thus results in an empty backward-looking center.

(62) B: a. en en en cinco días - hicimos dos mil setecientas citaciones

'in in five days, (we) did two thousand seven hundred notices'

A: b. ahá

'uh huh'

B: c. cuando él fue a la reunión de los abogados

'when he went to the meeting with the lawyers'

d. todos habían hecho cien, ciento veinte

'(they) had all done a hundred, a hundred and twenty'

e. no lo podían creer, viste

'(they) couldn't believe it, you know'

A: f. mirá, vos

'really?'

B: g. y, pero así también fue la gente que empezó a caer

'and, also that's how people realized'

h. imagínate, la mitad de la gente, toda caliente

'imagine, half the people, all mad'

i. porque le pedían impuestos que ya se, autos de hace treinta años que se transfirieron

'because (they) were being asked for taxes for cars that already, cars that had been transferred thirty years ago'

j. [pause] que no existen más, viste, ¡una goma! [A: { laugh }] - tremenda,

'that don't exist any more, you see, what a situation!'

k. entonces viste, el tipo vino calentón, así

'then, you see, the guy came back all excited, you know.'

Future research will be focused on the relationship between Centering and the discourse structure of the conversations, paying attention to conversational phenomena such as side sequences and turn-taking. I will also study the relationship between the local focus of attention (which Centering was devised to model) and the global structure of the conversations.

References

Alonso-Ovalle, Luis, Susana Fernández-Solera, Lyn Frazier and Charles Clifton. (2002). Null vs. overt pronouns and the topic-focus articulation in Spanish. *Rivista di Linguistica (Italian Journal of Linguistics), 14* (2), 151-170.

Amaral, Patrícia Matos and Scott A. Schwenter. (2005). Contrast and the (non-) occurrence of subject pronouns. In D. Eddington (Ed.), *Selected Proceedings of the 7th Hispanic Linguistics Symposium* (pp. 116-127). Somerville, MA: Cascadilla Press.

Ariel, Mira. (1988). Referring and accessibility. *Journal of Linguistics, 24*, 65-87.

Ariel, Mira. (1996). Referring expressions and the +/- coreference distinction. In T. Fretheim and J. K. Gundel (Eds.), *Reference and Referent Accessibility* (pp. 13-25). Amsterdam and Philadelphia: John Benjamins.

Asher, Nicholas. (1993). *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer.

Bolinger, Dwight L. (1979). Pronouns in discourse. In T. Givón (Ed.), *Syntax and Semantics, Vol. 12: Discourse and Syntax* (pp. 289-309). New York: Academic Press.

Brennan, Susan E. (1995). Centering attention in discourse. *Language and Cognitive Processes, 10* (2), 137-167.

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

Butt, John and Carmen Benjamin. (2000). *A New Reference Grammar of Modern Spanish*. Chicago: McGraw-Hill. (3rd edition).

Byron, Donna K. (2002). Resolving pronominal reference to abstract entities, *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL-02)* (pp. 80-87). Philadelphia, PA.

Byron, Donna K. and Amanda Stent. (1998). A preliminary model of Centering in dialog, *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.* Unpublished Ph.D. dissertation, University of Pennsylvania.

Carreiras, Manuel, Morton Ann Gernsbacher and Victor Villa. (1995). The advantage of first mention in Spanish. *Psychonomic Bulletin and Review, 2* (1), 124-129.

Clark, Herbert H. (1977). Bridging. In P. N. Johnson-Laird and P. C. Wason (Eds.), *Thinking: Readings in Cognitive Science* (pp. 411-420). Cambridge: Cambridge University Press.

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

Cornish, Francis. (2005). Degrees of indirectness: Two types of implicit referents and their retrieval via unaccented pronouns. In A. Branco, T. McEnery and R. Mitkov (Eds.), *Anaphora Processing: Linguistic, Cognitive and Computational Modelling* (pp. 199-220). Amsterdam and Philadelphia: John Benjamins.

Cote, Sharon. (1998). Ranking forward-looking centers. In M. A. Walker, A. K. Joshi and E. F. Prince (Eds.), *Centering Theory in Discourse* (pp. 55-69). Oxford: Clarendon.

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* (4), 543-565.

Di Eugenio, Barbara. (1990). Centering theory and the Italian pronominal system, *Proceedings of the 13th International Conference on Computational Linguistics (COLING-90)* (pp. 270-275). Helsinki, Finland.

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

Dimitriadis, Alexis. (1996). When pro-drop languages don't: Overt pronominal subjects and pragmatic inference. In L. M. Dobrin, K. Singer and L. McNair (Eds.), *CLS 32: Papers from the Main Session* (pp. 33-47). Chicago: Chicago Linguistics Society.

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 and Michael Strube. (1999). Resolving discourse deictic anaphora in dialogues, *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL-99)* (pp. 37-44). Bergen, Norway.

Edelsky, Carole. (1981). Who's got the floor? *Language in Society, 10*, 383-421.

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

Fais, Laurel. (2004). Inferable centers, Centering transitions and the notion of coherence. *Computational Linguistics, 30* (2), 119-150.

Fernández Soriano, Olga. (1999). El pronombre personal: Formas y distribuciones. Pronombres átonos y tónicos. In I. Bosque and V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 1: Sintaxis básica de las clases de palabras, pp. 1209-1273). Madrid: Espasa.

Geluykens, Ronald. (1994). *The Pragmatics of Discourse Anaphora in English: Evidence from Conversational Repair.* Berlin: Mouton de Gruyter.

Gernsbacher, Morton Ann and David J. Hargreaves. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language, 27* (6), 699-717.

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.

GNOME. (2000). *GNOME Project Final Report*. Edinburgh: University of Edinburgh.

Gordon, Peter C., Barbara J. Grosz and Laura A. Gilliom. (1993). Pronouns, names, and the Centering of attention in discourse. *Cognitive Science, 17* (3), 311-347.

Gordon, Peter C., Randall Hendrick, Kerry Ledoux and Chin Lung Yang. (1999). Processing of reference and the structure of language: An analysis of complex noun phrases. *Language and Cognitive Processes, 14* (4), 353-379.

Grice, H. Paul. (1975). Logic and conversation. In P. Cole and J. L. Morgan (Eds.), *Speech Acts. Syntax and Semantics, Volume 3* (pp. 41-58). New York: Academic Press.

Grosz, Barbara J., Aravind K. Joshi and Scott Weinstein. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics, 21* (2), 203-225.

Grosz, Barbara J. and Candace L. Sidner. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics, 12* (3), 175-204.

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

Hadic Zabala, Loreley and Maite Taboada. (2004). *Centering Theory in Spanish: Coding Manual*. Unpublished manuscript, Simon Fraser University.

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

Halliday, Michael A. K. and Ruqaiya Hasan. (1976). *Cohesion in English*. London: Longman.

Halliday, Michael A. K. and Christian M.I.M. Matthiessen. (2004). *An Introdution to Functional Grammar* (3rd ed.). London: Arnold.

Heap, David. (1998). Optimalizing Iberian clitic sequences. In J. Lema and E. Treviño (Eds.), *Theoretical Analyses on Romance Languages* (pp. 227-248). Amsterdam and Philadelphia: John Benjamins.

Hudson-D'Zmura, Susan and 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 and E. F. Prince (Eds.), *Centering Theory in Discourse* (pp. 199-226). Oxford: Clarendon.

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

Jackendoff, Ray. (1990). *Semantic Structure*. Cambridge, Mass: MIT Press.

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

Jefferson, Gail. (1972). Side sequences. In D. Sudnow (Ed.), *Studies in Social Interaction* (pp. 294-338). New York: Free Press.

Kameyama, Megumi. (1986). A property-sharing constraint in Centering, *Proceedings of the 24th Annual Meeting of Association for Computational Linguistics (ACL-86)* (pp. 200-206). New York, USA.

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

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.

Levin, Lori, Klaus Ries, Ann Thyme-Gobbel and Alon Lavie. (1999). Tagging of speech acts and dialogue games in Spanish Call Home, *Proceedings of ACL-99 Workshop on Discourse Tagging*. College Park, MD.

Luján, Marta. (1999). Expresión y omisión del pronombre personal. In I. Bosque and V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 1: Sintaxis básica de las clases de palabras, pp. 1275-1315). Madrid: Espasa.

Mendikoetxea, Amaya. (1999). Construcciones con *se*: medias, pasivas e impersonales. In I. Bosque and V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 2: Las construcciones sintácticas fundamentales; Relaciones temporales, aspectuales y modales, pp. 1631-1722). Madrid: Espasa.

Miltsakaki, Eleni and Karen Kukich. (2004). Evaluation of text coherence for electronic essay scoring systems. *Natural Language Engineering, 10* (1), 25-55.

Pesetsky, David. (1987). *Wh*-in-situ: Movement and unselective binding. In E. Reuland and A. G. B. ter Meulen (Eds.), *The Representation of (In)definiteness* (pp. 98-129). Cambridge, MA: MIT Press.

Pesetsky, David. (2000). *Phrasal Movement and Its Kin*. Cambridge, MA: MIT Press.

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

Poesio, Massimo, Rosemary Stevenson, Barbara Di Eugenio and Janet Hitzeman. (2004). Centering: A parametric theory and its instantiations. *Computational Linguistics, 30* (3), 309-363.

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

Roberts, Craige. (1998). The place of Centering in a general theory of anaphora resolution. In M. A. Walker, A. K. Joshi and E. F. Prince (Eds.), *Centering Theory in Discourse* (pp. 359-399). Oxford: Clarendon.

Schegloff, Emmanuel, Gail Jefferson and Harvey Sacks. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*, 361-382.

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

Sharp, Randy. (2005). A unified treatment of Spanish *se*. In A. Branco, T. McEnery and R. Mitkov (Eds.), *Anaphora Processing: Linguistic, Cognitive and Computational Modelling* (pp. 113-136). Amsterdam and Philadelphia: John Benjamins.

Sidner, Candace L. (1983). Focusing in the comprehension of definite anaphora. In M. Brady and R. C. Berwick (Eds.), *Computational Models of Discourse* (pp. 267-330). Cambridge, Mass: MIT Press.

Stevenson, Rosemary, Rosalind A. Crawley and David Kleinman. (1994). Thematic roles, focus and the representation of actions. *Language and Cognitive Processes, 9*, 519-548.

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 and Udo Hahn. (1999). Functional Centering: Grounding referential coherence in information structure. *Computational Linguistics, 25* (3), 309-344.

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

Taboada, Maite. (2000). Cohesion as a measure in generic analysis. In A. Melby and A. Lommel (Eds.), *The 26th LACUS Forum* (pp. 35-49). Chapel Hill, N.C.: The Linguistic Association of Canada and the United States.

Taboada, Maite. (2002a). Centering and pronominal reference: In dialogue, in Spanish, *Proceedings 6th Workshop on the Semantics and Pragmatics of Dialog, EDILOG* (pp. 177-184).

Taboada, Maite. (2002b). Foco y pronominalización en la lengua hablada: Una primera aproximación. *Documentos de español actual, 3-4*, 173-200.

Taboada, Maite. (2004). *Building Coherence and Cohesion: Task-Oriented Dialogue in English and Spanish*. Amsterdam and Philadelphia: John Benjamins.

Taboada, Maite and Loreley Hadic Zabala. (2005). *What are the units of discourse structure? Segmenting discourse within Centering Theory*. Unpublished manuscript, Simon Fraser University.

Tetreault, Joel R. (2001). A corpus-based evaluation of Centering and pronoun resolution. *Computational Linguistics, 27* (4), 507-520.

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

Walker, Marilyn A. (1998). Centering, anaphora resolution, and discourse structure. In M. A. Walker, A. K. Joshi and E. F. Prince (Eds.), *Centering Theory in Discourse* (pp. 401-435). Oxford: Clarendon.

Walker, Marilyn A., Masayo Iida and Sharon Cote. (1994). Japanese discourse and the process of Centering. *Computational Linguistics, 20* (2), 193-232.

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

Walker, Marilyn A. and Ellen F. Prince. (1996). A bilateral approach to givenness: A hearer-status algorithm and a Centering algorithm. In J. K. Gundel and T. Fretheim (Eds.), *Reference and Referent Accessibility* (pp. 291-306).

Wanner, Dieter. (1994). El orden de los clíticos agrupados en castellano. *Thesaurus, 49* (1), 1-57.

Webber, Bonnie Lynn. (1981). Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes, 6* (2), 107-135.

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.

Zaenen, Annie, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M. Catherine O'Connor and Tom Wasow. (2004). Animacy encoding in English: Why and how. In B. Webber and D. K. Byron (Eds.), *Proceedings of ACL-2004 Workshop on Discourse Annotation* (pp. 118-125). Barcelona, Spain.