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 (

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}

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, whethe

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

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*

(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):

 list is

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

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

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.

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?

     B :   b .   m ó d e m ,   l o s   t e n e m o s
                                                 '   M

     c. pero no los instalamos todavía
                                               '   b   u
(3   7   )   G   r   a   m   m   a
     a. Cf:
                                        C   b   :
          b .   C   f   :
             C b– Transition: RETAIN
     c. Cf:

---

19 Thanks to Laurie Fais
[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

          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

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 Center

context, but the inference seemed a bit far-fetched, and I decided not to establish it. That is the

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

(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

'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), or10.w2

(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.'

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

(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,
      your Dad 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 nego

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

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’

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. Inss-ramela A. (22.825 04 s

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 9Giv

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*