

Centering Theory in Spanish: Coding Manual*

Loreley Hadic Zabala and Maite Taboada

lmhadic@sfu.ca, mtaboada@sfu.ca

Simon Fraser University

Current version: June 6, 2006

0. Introduction

This is a manual for coding Centering Theory (Grosz et al., 1995) in Spanish. The manual is still under revision. The coding is being done on two sets of corpora:

- ISL corpus. A set of task-oriented dialogues in which participants try to find a date where they can meet. Distributed by the Interactive Systems Lab at Carnegie Mellon University. Transcription conventions for this corpus can be found in Appendix A.
- CallHome corpus. Spontaneous telephone conversations, distributed by the Linguistics Data Consortium at the University of Pennsylvania. Information about this corpus can be obtained from the LDC.

This manual provides guidelines for how to segment discourse (Section 1), what to include in the list of forward-looking centers (Section 2), and how to rank the list (Section 3). In Section 4, we list some unresolved issues.

1. Utterance segmentation

1.1 Utterance

In this section, we discuss how to segment discourse into utterances. Besides general segmentation of coordinated and subordinated clauses, we discuss how to treat some spoken language phenomena, such as false starts.

In general, an utterance *U* is a tensed clause. Because we are analyzing telephone conversations, a turn may be a clause or it may be not. For those cases in which the turn is not a clause, a turn is considered an utterance if it contains entities.

The first pass in segmentation is to break the speech into intonation units. For the ISL corpus, an utterance *U* is defined as an intonation unit marked by either {period}, {quest} or {seos} (see Appendix A for details on transcription). Note that {comma}, unless it is followed by {seos}, does not define an utterance.

In the example below, (1c.) corresponds to the beginning of a turn by a different speaker. However, even though (1c.) is not a tensed clause, it is treated as an utterance because it contains entities, it is followed by {comma} {seos}, and it does not seem to belong to the following utterance.

* This work is supported by the Ministry of Science and Technology of Spain and the Xunta de Galicia, under project MCYT-FEDER BFF2002-02441/XUGA -PGIDIT03PXIC20403PN, and by Simon Fraser University, under a SSHRC grant and a Discovery Park grant.

- 1 a. fvgc: así que si Ø te viene bien
so that if nullpro:3SG OBJ:2SG go:3SG.PRES well
- de diez a doce {comma}
from ten to twelve
- ‘So if (it) is good for you from ten to twelve’
Cf: fsm (te), 10-12
Cb: 0
- b.este Ø est-á bien {period} #key_click# {seos}
eh nullpro:3SG be -3SG:PRES well
- ‘then, (it) is good’
Cf: 10-12 (zero)
Cb: 10-12
Transition: CONTINUE
- c. fsm: perfecto {period} {seos} diez a doce el veintitrés {comma}{seos}**
perfect ten to twelve the twenty-three
- ‘Perfect. (From) ten to twelve (on) the twenty-third’
Cf: 10-12, 23
Cb: 10-12
Transition: CONTINUE
- d. y bueno {seos}/ls/ Ø te ve -ré en tu
and well nullpro:1SG OBJ:2SG see-1SG:FUT in POSS:2SG
- oficina entonces {period} {seos}
office then
- ‘Well, (I) will see you in your office then.’
Cf: fsm (nullpro), fvgc (te, tu), oficina, 10-12 (entonces)
Cb: 10-12
Transition: RETAIN

For the CallHome corpus, an utterance U consists of a Dialogue Act </DA>. Dialogue Acts were coded in LDC corpus LDC2001T61.

- 2 a. B: Sí, tío. ¿Y chic-o -s? </DA>
Yes uncle And kid -MASC -PL
- ‘Yes, uncle. And kids?’
Cf: A, chicos
Cb: 0
- b. A: Todavía no </DA>
yet not
- ‘Not yet.’
Cf: A, chicos
Cb: A
Transition: CONTINUE

Following Byron and Stent (1998:5), “empty utterances”, that is, utterances that contain no discourse entities, are attached to their preceding or following utterance, according to context. In Example (3), *bueno* is attached to the following utterance.

- 3 fcb_a_08_02: /h#/ **bueno** {period} {seos} el ocho yo pued-o {comma}
 well the eight I can -1SG:PRES
 pause /s/ a part -ir de l -as dos {period}
 at start -INF of the -FEM:PL two
 ‘Well, on the eighth I can after (lit. ‘starting at’) two.’

This applies to empty utterances across turns as well, so that backchannels are ignored for Centering purposes. In (4), the backchannel *sí* by speaker B is ignored as an utterance.

- 4 A: Ø Me levant-o a l -as siete --
 nullpro:1SG 1SG get.up-1SG:PRES at the-FEM:PL seven
 ‘(I) get up at seven- -
 Cf: A, 7
 Cb: A
 Transition: CONTINUE
- B: Sí.**
 Yes
 ‘Yes’
- A: Ø empiez-o l -as clase-s de ocho
 Nullpro:1SG start -1SG:PRES the-FEM:PL class-FEM:PL from eight
 to
 a nueve cuarenta --
 to nine forty
 ‘(I) have class from eight till nine forty’
 Cf: A, clases, 8-9:40
 Cb: A
 Transition: CONTINUE

1.2 Discourse segmentation: conjuncts and adjuncts¹

If intonation units are not marked in the discourse, we will follow Kameyama (1998). The hypothesis proposed by Kameyama is that clauses within a (grammatically complete) sentence are center-updating units, i.e., Centering should be considered not only across sentences, but also within sentences. She proposes two models of intrasentential Centering: (i) Sequential intrasentential Centering, where each unit (clause, conjunct or adjunct) is processed at a time, resulting in a flat discourse tree, with one Centering state at a time; (ii) Hierarchical intrasentential Centering, where multiple Centering states are possible, at different levels of embedding. Sequential intrasentential Centering applies to tensed and tenseless conjuncts and tensed adjuncts. Hierarchical intrasentential Centering applies to reported speech complements, tensed non-report complements and relative clauses.

¹ Conjuncts are expressions that are joined together by coordinating conjunctions (and/or/but) or subordinating conjunctions (Radford 1997:499-500).
 Adjuncts are expressions that provide additional information about the relevant activity/event, such as location, time and manner in which it took place (Radford 1997:142).

We are, for the time being, considering the first model of intrasentential Centering as our general model of Centering. That is, each clausal unit (segmented as described below) is a Centering unit. We believe this is the most appropriate model for spoken discourse. Exceptions are those mentioned by Kameyama: reported speech and non-report complements, where the reported part is embedded in the same Centering unit as the reporting unit (see below). These are to be processed differently: the embedded part becomes a segment and undergoes Centering analysis, but is not considered an update unit for the following clause. This is the approach taken by Suri and McCoy (1994) for processing main-subordinate clauses pairs (“X because Y”). We do not believe that approach is appropriate for tensed subordinate clauses (Taboada and Hadic Zabala, 2004), but do follow Kameyama in applying it to reported speech and non-report complements.

In order to establish which clausal units will be considered as Centering units, we follow Kameyama’s guidelines for sentence segmentation. The following are all taken from Kameyama (1998:104-108), including the names for the hypotheses. We illustrate them with examples from our corpora.

1.2.1 Tensed conjuncts: Tensed clausal conjuncts break up into a sequence of utterances (Tensed conjunct hypothesis, TConj). In (5), the two coordinated sentences are separated into two utterances. Note that (5b) contains an embedded non-finite clause, not separated into its own utterance.

- 5 a. así <[n]> <[n]> ∅ combin-amos bien /ls/
 so.that nullpro:1PL plan -1PL:PRES well
 ‘So that (we) can plan it well’
 Cf: nosotras (nullpro)
 Cb: 0
- b. y Æ **qued -amos** **para** **ver** =nos {period} {seos}
 and nullpro:1PL arrange-1PL:PRES to see:INF =1PL:RFL
 ‘and (we) can arrange to see each other’
 Cf: nosotras (nullpro, -nos)
 Cb: nosotras
 Transition: CONTINUE

TConj also applies to tensed **subordinate** conjuncts. (6a.), (6b.) and (6c.) below are three tensed subordinate clauses belonging to the main clause in (6d.). In our analysis, each tensed subordinate clause is treated as a separate utterance.

- 6 a.A: **si** Æ **te** **enganch -ás** **con** **es -o -s,**
 if nullpro:2SG RFL:2SG hook.up -2SG:PRES with that-MASC-PL
 ‘If (you) hook up with those’
 Cf: B (nullpro, te), Compuserve (esos)
 Cb: Compuserve
 Transition: ROUGH SHIFT

b. **y** \AA **ten -és** **acceso,**
 and nullpro:2SG have-2SG:PRES access

‘And (you) have access’

Cf: B (nullpro), internet (acceso)

Cb: B

Transition: SMOOTH SHIFT

c. **yo** **ten -go** **también** **acceso,**
 I have-1SG:PRES too access

‘I also have access’

Cf: A (yo), internet (acceso)

Cb: internet

Transition: ROUGH SHIFT

d. \emptyset nos pode -mos comunic -ar por por
 nullpro:1PL RFL:1PL be.able-1PL:PRES communicate-INF by by

 correo, así
 mail so

‘(We) can communicate by e-mail’

Cf: nosotros (nullpro, nos), internet (correo)

Cb: internet

Transition: RETAIN

1.2.2 Conjunct parallelism: Adjacent conjuncts, whether tensed or tenseless, induce parallelism (Conjunct parallelism hypothesis, Cpara). This is particularly relevant for cases of subject ellipsis or gapping.

7 a. "A" Kiko una vez se gan-ó veinte,
 Kiko one time refl:3sg win-3sg:pret twenty

‘Kiko once won twenty (dollars)’

Cf: Kiko, premio (veinte)

Cb: premio

Transition: RETAIN

b. **Pato,** **también.** </DA>
 Pato too

‘Pato (did so) too.’

Cf: Pato, premio

Cb: premio

Transition: RETAIN

8 a. \emptyset teng -o dos, dos \emptyset teng -o, sí
 nullpro:1SG have-1SG:PRES two two nullpro:1SG have-1SG:PRES yes

‘(I) have two, two (I) have, yes’

Cf: B (nullpro), laburos (dos)

Cb: B

Transition: CONTINUE

i. \emptyset **no** **me** **bronc -a.** </DA>
 nullpro:3SG not OBJ:1SG bother-3SG:PRES

‘(It) doesn’t bother me’

Cf: maestra (me), \emptyset (nullpro)

Cb: maestra

Transition (10b to 10b.i): SMOOTH SHIFT

ii. **El** **lunes** \emptyset **te** **voy** **a** **hac-er**
 the Monday nullpro:1SG 2SG:OBJ go:1SG:PRES to do -INF

tu **examen.**
 POSS:2SG exam

‘Monday (I) will give you your exam’

Cf: B (te), maestra (nullpro), exam, Monday

Cb: maestra

Transition (10b.i to 10b.ii): RETAIN

1.4 Nonreport complements

1.4.1 Tensed nonreport complements: Tensed clausal nonreport complements create embedded discourse segments (Clausal complement hypothesis, Comp). Even though the tensed complement in (13a.i.) constitutes a centering unit, it does not update the center for the following utterance. Thus, for (13b.) the previous utterance is (13a.) and not (13a.i).

13 a. B: **sí,** **yo** **lo** **quier-o** **convenc -er**
 yes I OBJ:3SG:MASC want-1SG:PRES convince-INF

‘I want to convince him’

Cf: B, tipo

Cb: B

Transition: SMOOTH SHIFT

a.i **que** \emptyset **me** **haga** **un** **préstamo**
 that nullpro:3SG OBJ:1SG do:3SG:SUBJ a:MASC loan

a **mí**
 to OBJ:1SG

‘that (he) should give me a loan’

Cf: tipo(nullpro), B, préstamo

Cb: B

Transition: RETAIN

b. **ahí** **si** **que** **se** **me** **acab-a** **la**
 there yes that RFL OBJ:1SG end -3SG:PRES the:SG:FEM

‘then (something) would be over for me’

Cf: B, la?

Cb: B

Transition (13a to 13b): CONTINUE

As noted by Thompson (2002:134), certain complement-taking-predicates (CTP) “do not constitute the speakers’ interactional agenda, but are instead functioning to convey the speakers’ epistemic, evidential, or evaluative stance towards the issue or claim at hand.” In English, the most common examples of such CTPs are *think*, *know*, and *guess*, and

they tend to occur with 1st person subjects (for comprehensive list, see Thompson 2002: 138). Following Thompson (2002: 136), these CTPs and their clausal complements will be analyzed as monoclausal utterances. In other words, the clausal complements of epistemic, evidential or evaluative CTPs do not constitute embedded segments. CTPs express the epistemic/evidential/evaluative stance of the speaker towards the information contained in the complement clause, and could be substituted by modals or adverbs (Thompson, 2002: 132). The analysis of these clauses is a flat analysis, i.e., as if there was no embedding. The subject of the CTP is typically the first entity in the Cf list.

Examples (14) and (15) illustrate this type of construction. In (14d.) the verb *creo* ‘believe’ creates an epistemic frame for the clause that follows. It is the speaker’s belief that his friend ended her relationship with her boyfriend in England.

- 14 a."B" una amiga que dejó la escuela, le entró la locura y se fue con su novio que estudia medicina en [PAUSE] el Medical College o algo así, de Inglaterra, y se largó con él
 ‘A friend who left school, went crazy and left with her boyfriend who studies medicine at the English Medical College, or something like that, in England, and she went with him.’
- b."A" Mmm </DA>
- c."B" y este, </DA> y Ø se fue a
 and uh and nullpro:3SG 3SG:RFL go: 3SG:PAST to
 la aventura,
 the: FEM:SG adventure
 ‘And, uh, she went looking for adventure’
- d. después Ø cre -o que Ø
 afterwards nullpro:1SG believe-1SG:PRES that nullpro:3SG
 l -o cort-ó por allá
 OBJ-MASC:SG cut -3SG:PAST for there
 ‘Afterwards I think she ended it over there’
 Cf: B (nullpro), she (nullpro), it (=relationship), there (=England)

Example (15) further emphasizes the parenthetical nature of CTPs. ‘Yo creo’ does not occur in the canonical position, i.e. preceding the complement clause, but within the complement clause. As indicated by Thompson (2002: 134) epistemic parentheticals may “float away from CTP position.”

- 15 a. "A" Y entonces. Pero cómo vas \emptyset a
 And then But how go: 2SG:PRES nullpro:2SG to
 hacer tesis, ahora, entonces.
 do.INF thesis now then
 'Then, how are you going to do your thesis now?'
- b. "B" Pues, \emptyset I -a voy a
 Well nullpro:1SG OBJ-FEM:SG go:1SG:PRES to
 hacer **yo cre -o,** en el
 do.INF 1SG believe-1SG:PRES in the:MASC:SG
 Centro de Ecología -- </DA>
 Center of Ecology
 'Well, I'll do it, I think, at the Ecology Centre.'
 Cf: B (nullpro), thesis, Ecology Centre

1.4.2 Tenseless nonreport complements: Tenseless clausal complements belong to the same utterance units as their superordinate clauses (Tenseless complement hypothesis, TlessComp). Example (16) is a single utterance, with an embedded tenseless verb, *traer*.

- 16 por favor \emptyset no te olvid -es **de**
 please nullpro:2SG not 2SG:RFL forget-2SG:PRES.IMPR of
- traer tod-os I -os legajo-s** {period}
 bring.INF all -MASC:PL the- MASC:PL file -PL
 'Please, don't forget to bring all the files.'
 Cf: fmgl (nullpro, te), legajos

1.5 Relative clauses

Relative clauses create embedded centering units. Their entities, however, may be accessible to superordinate clauses (p.108). As indicated by Poesio et al. (to appear: 31), this would mean that relative clauses have a special status: at the local level of discourse, they update the local focus and at the global level of discourse, they merge with the superordinate clause. Because this approach would lead to a radical change in the Centering framework, Poesio et al. propose to treat relative clauses as separate utterances, although as embedded utterances and not as center-updating utterances. We follow this approach. In Example (17), the relative clause in (17b.i.) is processed as an utterance, but does not serve as previous utterance for (17c.)—i.e., it is not a center-updating unit.

- 17 a. y después \emptyset se fue a Miami
 and then nullpro:3SG 3SG:RFL go:3SG:PRET to Miami
 'And then (he) went to Miami'
 Cf: tipo (nullpro), Miami
 Cb: tipo
 Transition: CONTINUE

b. Ø Se compr-ó algun-as cosa -s de est -a -s
 nullpro:3SG 3SG:RFL buy -3SG:PRET some-FEM:PL thing-PL of this-FEM-PL
 ‘(He) bought some of these things’
 Cf: tipo (nullpro), cosas
 Cb: tipo
 Transition: CONTINUE

i. **que** Æ **le** **pusimos** **a** **l -a**
 that nullpro:1PL OBJ:SG put:PRET:1PL to the-FEM:SG

máquina
 machine
 ‘that (we) put in the computer’
 Cf: nosotros (nullpro), cosas (que), máquina
 Cb: cosas
 Transition: ROUGH SHIFT

c. él siempre va de compras qué sé yo
 he always go.3SG:PRES of shopping what know: 1SG:PRES I
 ‘He always goes shopping, what do I know’
 Cf: tipo (él), compras, yo
 Cb: tipo
 Transition (15b to 15c): CONTINUE

While headed relative clauses (as in 17b.i.) create embedded centering units, headless relative clauses (18) and certain idiomatic expressions (usually of time) containing relative clauses (19) do not, and are thus analyzed in conjunction with their superordinate clauses.

18 A: el que se va, el el
 the:MASC:SG that 3SG:RFL go:3SG:PRES the:MASC:SG the:MASC:SG

 que se va alrededor del veinte
 that 3SG:RFL go:3SG:PRES around of:the:MASC:SG twentieth

 es Pablo Lates eh
 be:3SG:PRES Pablo Lates eh
 ‘the one who leaves around the 20th is Pablo Lates, eh’
 Cf: Pablo Lates, 20th

19 yo tengo **la** **semana que** **viene** una
 I have: 1SG:PRES the:FEM:SG week that come:3SG:PRES a:FEM:SG

 reunión en IBM
 meeting in IBM
 ‘I have next week (lit. the week that comes) a meeting at IBM’
 Cf: I, meeting, IBM, next week

1.6 False starts

Speakers may correct themselves or abandon utterances altogether. Following Eckert and Strube (1999), self-corrected speech (false starts and speech repairs) are taken into account, since they may introduce a new entity in the discourse. With respect to segmentation, false starts belong to the same utterance as the self-repaired speech that

follows. In terms of the Cf list, the ranking of the entities in the false start with respect to the entities in the repaired speech proceeds linearly. Note however, that only the false starts that contain entities are taken into account. This is illustrated in example (20). In (20a.), *te*, a pronoun referring to the addressee, becomes part of the Cf list. In (20b), there are no entities in the false start (marked with angled brackets), and therefore there is nothing to include in the Cf list.

- 20 a. fmcs_01_11: *pause* bueno {period} {seos} < te {seos} > /mm/ entonces
 well 2SG:OBJ then
- ∅ qued -amos así {period} {seos}
 nullpro:1PL arrange-1PL:PRES so/like.this
 ‘Well, <you> then (we) agree on this’
 Cf: fmg1 (te), nosotras
 Cb: 0
 Transition: NONE
- b. por favor no ∅ te olvid -es
 please not nullpro.2SG 2SG:RFL forget-2SG:PRES
- de tra -er tod-os l -os legajo-s {period}
 of bring-INF all -MASC:PL the-MASC:PL file -PL
- /h#/ < para pod -er este > para ten -er
 to be.able-INF eh to have-INF
- tod-a l -a información a mano
 all -FEM:SG the-FEM:SG information at hand
- {period}{seos}
 ‘Please, don’t forget to bring the files **in order to be able eh** in order to have all the information at hand.’
 Cf: fmg1 (nullpro, te), legajos, información
 Cb: fmg1
 Transition: CONTINUE

2. Realization

When computing the Cb (backward-looking center) of an utterance, we allow for indirect realization of entities. In the corpora studied, constraining realization only to the entities that have been explicitly mentioned in the utterance led to a larger number of transitions with an empty Cb. Following Halliday and Hasan’s (1976) definition of lexical cohesion, the following types of indirect realization are allowed.

2.1 General noun

- 21 a. "B" sí, y quien es **Mónica Martínez,** no? </DA>
 yes and who be:3SG:PRES Mónica Martínez not
 ‘Yes, and who is **Mónica Martínez**, right?’
 Cf: Mónica Martínez
 Cb: 0
 Transition: NONE

b. nadie l -a co- </DA>
 nobody OBJ-FEM:SG knw-
 ‘Nobody (knows) her’
 Cf: nadie, Mónica Martínez (la)
 Cb: Mónica Martínez
 Transition: RETAIN

c. "A" un-a much- **un-a** **muchacha**
 a -FEM:SG gir- a -FEM:SG girl
 ‘A gir- a girl’
 Cf: Mónica Martínez (muchacha)
 Cb: Mónica Martínez
 Transition: CONTINUE

 que nac -ió en Camiri [PAUSE] Cochabamba, Bolivia </DA>
 that be.born-3SG:PRET in Camiri Cochabamba Bolivia
 ‘who was born in Camiri, Cochabamba, Bolivia.’
 Cf: Mónica Martínez (que), Camiri, Cochabamba, Bolivia
 Cb: Mónica Martínez
 Transition: CONTINUE

2.2 Synonyms and near synonyms (when they have the same reference)

In examples (22a.) and (22c.) below, the words *picture* and *icon* have the same reference and are used as synonyms.

22 a. B: if someone could send me the %um **the blessed virgin picture**
 Cf: someone, B (me), picture
 Cb: B
 Transition: ROUGH SHIFT

 that I have in my room
 Cf: B (I), picture (that), B (my), room
 Cb: B
 Transition: CONTINUE

b. A: okay

c. B: **the icon** that's next to that gold %uh cross
 Cf: icon, gold cross
 Cb: icon
 Transition: SMOOTH SHIFT

 that I have
 Cf: B (I), cross (that)
 Cb: cross
 Transition: ROUGH SHIFT

2.3 Superordinate

- 23 a. "B" Sí, </DA> además Ø no te dijeron que tipo
 yes also nullpro:3PL not 2SG:OBJ say:PRET:3PL what type
 de **ganado**, </DA>
 of cattle
 'And also, (they) didn't tell you what type of **animal**.'
 Cf: 3pl (nullpro), A (te), ganado
 Cb: ganado
 Transition: RETAIN
- b. a.lo.mejor Ø son **topo-s**, o -- </DA>
 maybe nullpro:3PL be:3PL:PRES mole-PL or
 'Maybe (they) are **moles**.'
 Cf: ganado (topos)
 Cb: ganado
 Transition: CONTINUE

2.4 Inclusive relation

- 24 a. "A" Y l -os, mijja, y l -os
 and the-MASC:PL my.daughter and the-MASC:PL
niñ-it -os qué tal est-án. </DA>
 kid -DIM-MASC:PL what ¿? be -3PL:PRES
 'And the, dear, and **the kids** how are they?'
 Cf: niños
 Cb: 0
 Transition: NONE
- b. "B" Bien, </DA> **Samuel** ayer se ca-yó en
 well Samuel yesterday 3SG:RFL fall-3SG:PRET in
 l -a pisci- afuera de l -a piscina, </DA>
 the-FEM:SG swim- outside of the-FEM:SG swimmingpool
 'Well, yesterday **Samuel** fell in the swim- outside the swimmingpool.'
 Cf: Samuel (one of niños), piscina
 Cb: Samuel
 Transition: CONTINUE

2.5 Part – whole

- 25 a. I mean there was **trees** down
 Cf: trees
 Cb: 0
 Transition: NONE
- b. there was **branches** all over
 Cf: trees (branches)
 Cb: trees
 Transition: CONTINUE

3. Cf –Ranking

3.1 Ranking criterion

The most important aspect of adapting Centering Theory to a new language is to determine the ordering of the Cf list, what Cote (1998) calls the *Cf template* for a language.

We 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 (26), or as null pronouns.

26	como	vos		me	has		dicho	en
	like	2SG:SUBJ		1SG:OBJ	have: 2SG:PRES		say:PASTPART	in
		un -a	de	tu	-s	carta-s,		¿no? </DA>
		one-FEM:SG	of	POSS:2SG-PL		letter- PLno		

‘Like **you** have told **me** in one of your letters, right?’
 Cf: A (vos), me (B), carta

However, there are two other criteria that play a role in the Cf ordering in Spanish: empathy and animacy.

Following DiEugenio (1998), we 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.”

Empathy in Spanish, from our point of view, is reflected in the experiencer in psychological verbs, where the point of view taken is that of the experiencer, regardless of whether it is the subject or not (e.g., “it seems to **me**”, “**I** think”, and the like). In (27) 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: the PP “a mí”, plus the clitic “me”. In Example (28), the clitic “me” refers to the speaker, for whom Thursday is a better date. However, the point of view criterion need not apply to the speaker only. In (29), the point of view is that of the interlocutor.

A number of verbs in Spanish follow this pattern (“me conviene”, “me viene mejor”, “se me hace que”).

27	a	mí	me	parece-Ø		también {comma} /crky/
	to	1SG	1SG	seem -3SG:PRES		too
		bueno	de	hac-er	un -a	reunión {comma} {seos}
		ok	of	do -INF	a -FEM:SG	meeting

‘To **me**, it also seems ok to have a meeting’
 Cf: mphb (mí, me), [hacer una reunión], reunión
 Cb: mphb

- 28 /h#/ **me** viene mejor el jueves {comma} {seos}
 1SG come:3SG:PRES better the:MASC:SG Thursday
 ‘Thursday is better for **me**’
 Cf: mphb (me), jueves
 Cb: 0
- 29 /h#/ *pause* este /ls/ qué.tal para **tú** {comma} *pause*
 so how for 2SG
 del quince a -l diecinueve {period} {seos}
 from.the:MASC:SG fifteenth to-the:MASC:SG nineteenth
 ‘How is it for **you** from the fifteenth to the nineteenth?’
 Cf: meba (tú), del 15 al 19
 Cb: 0

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 (30), the experiencer is in a prepositional phrase (*para mí*, ‘for me’). We believe the experiencer should be ranked higher than either the clause as a whole that has the function of subject (*juntarme con vos ese día*, ‘to get together with you that day’), , or any of the entities included in that clause.

- 30 así que para **mí** ser-ía imposible
 so that for 1SG be -PRES.COND impossible
 junt-ar -me con vos /h#/ /eh/ es -e día /h#/
 join-INF -1SG with 2SG that-MASC:SG day
 {period} {seos}
 ‘So it would be impossible for **me** to join you that day’
 Cf: fmg1 (mí), [juntarme...], fmcs (-me, vos), ese día (miércoles 17)
 Cb: fmg1

Animacy is a relevant feature in the ordering of clitics and reflexive pronouns that refer to participants in the discourse. These have two characteristics that make them candidates for a higher ranking: (i) they convey empathy and (ii) they are often placed before the verb, linearly before non-animate direct objects. In (31), the indirect clitic *me* refers to the speaker, whereas the direct *la* refers to a thing, the computer. In (32), *se* refers to a third person, the speaker’s sister. In both cases, the animate indirect clitic is first in surface position.

- 31 A: yo **me** **l** -a compr-é usad-a
 I 1SG OBJ-FEM:SG buy -1SG:PRET used-FEM:SG
 ‘I bought a used one for me’
 Cf: A (yo, me), computadora (la)
 Cb: computadora

- 32 "A" ; Y que Ø se I -o d -an! </DA>
 and that nullpro:3PL 3SG OBJ-MASC:3SG give-3PL:PRES
 'And **it** was given to **her**'
 Cf: hermana (se), programa (lo), imp -3pl (nullpro)
 Cb: hermana

Subjects still take precedence in the Cf list in most other cases (i.e., when they are not clausal, and when there are no experiencers). In Example (33), the subject *maestra* is higher in the Cf list than the clitic *me*. This is in turn higher than the direct object *mi examen*. The ordering of the direct object NP follows the possessive ordering (see 3.2).

- 33 Un-a maestra este, me tuvo que ven -ir a
 a -FEM:SG teacher eh OBJ:1SG have:3SG:PRET that come-INF to
 hac-er mi últim-o examen aquí. </DA>
 do -INF POSS:1SG last -MASC:SG exam here
 'A teacher had to come and give me my exam here.'
 Cf: maestra, B (me), B (mi), examen, aquí
 Cb: 0

Accordingly, the elements of the Cf list follow this order:

Empathy > Subj > Animate IObj > DObj > etc > Subj > Animate IObj > DObj > etc²
 Main clause Subordinate clauses

3.2 Possessives

Following DiEugenio (1998), we rank the possessor before the possessed, if the possessed is inanimate, and the possessor after the possessed, if the possessed is animate.

- 34 Un-a maestra este, me tuvo que ven -ir a
 a -FEM:SG teacher eh OBJ:1SG have:3SG:PRET that come-INF to
 hac-er mi últim-o examen aquí. </DA>
 do -INF POSS:1SG last -MASC:SG exam here
 'A teacher had to come and give me my exam here.'
 Cf: B (me), maestra, B (mi), examen, aquí
 Cb: 0

- 35 Mi hermana solicit -ó un programa
 POSS:1SG sister apply.for-1SG:PRET a:MASC:SG program
 de arqueología y antropología en Grecia. </DA>
 of archeology and anthropology in Greece
 'My sister applied for a program in archeology and anthropology in Greece.'
 Cf: hermana, A (mi), programa, Grecia
 Cb:0

This also applies to noun phrases with a PP modifier headed by "of" (*de* in Spanish). In most of those constructions, the meaning is that of a genitive (*las cartas de Marta =*

² This Cf template is slightly different from that proposed in previous publications (Taboada 2002a, 2002b).

Marta's letters)³. Thus, in Example (36), *una de Marta* refers to one (letter) from Marta. Since Marta is animate, it is ranked higher.

36 Y un -a de Marta.
 And one-FEM:SG of Marta
 'And one (letter) from Marta'
 Cf: Marta, 1 (=carta)
 Cb: carta

3.3 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 play. In (37b.), the wh-pronoun *qué* ('what') is included in the Cf list. It functions as direct object in the clause, and since there are no other entities, it is ranked highest. In (37c.), the NP 'environmental science' links back to the wh-pronoun and answers the question raised in (37b.): 'environmental science' is **what** the teacher is doing at the University of South Wales.

37 a. "B" Ø se va a l -a Universidad
 nullpro:3SG 3SG go:3SG:PRES to the-FEM:SG university

de Gales, del Sur,
 of Wales of-the.MASC:SG South

'(She) is going to the University of South Wales'

Cf: maestra(nullpro), universidad

Cb: maestra

Transition: RETAIN

i. donde estudió Sarucán, también. </DA>
 where study-3SG:PRET Sarucán too

'where Sarucán studied as well.'

Cf: Sarucán, universidad (=donde)

Cb: universidad

Transition: ROUGH SHIFT

b. "A" A hac-er qué. </DA>
 to do -INF what.OBJ

'To do **what**?'

Cf: maestra, **qué** (OBJ)

Cb: maestra

Transition: CONTINUE

c. "B" Este. </DA> A hac-er **ecología.** </DA>
 eh to do -INF environmental science

'Eh, to study **environmental science.**'

Cf: maestra, **ecología** (**qué**)

Cb: maestra

Transition: CONTINUE

³ Contra Walker and Prince's (1996) Complex NP Assumption, which ranks complex NPs (such as an NP with a possessive) in linear order, left-to-right. Since we are considering animacy as a relevant feature, we preferred to follow Di Eugenio's ranking for possessives, and to expand it to other complex NPs.

3.4 Relative pronouns

Relative pronouns should be ranked according to the role of the pronoun in the relative clause Subj>obj>etc., for the purpose of computing the Cf list. However, Poesio et al. (to appear) have shown that relative pronouns are not affected by Rule 1 of Centering, i.e. the Cb need not be a pronoun when a non-Cb relative pronoun is present. In (38), the relative pronoun *que* ('that') realizes the direct object of the relative clause and is thus ranked below the subject.

38	pero but	es be:3SG:PRES	que that	∅ nullpro:1SG	no not	he have:1SG:PRES	pod -ido be.able-PASTPART
		junt -ar gather-INF	l -a the-FEM:SG	plata money			
	'But (I) have not been able to gather the money'						
	Cf: A (nullpro), plata						
	Cb: A						
	Transition: CONTINUE						
	que that	∅ nullpro:1SG	necesit-o need -1SG:PRES	para to	pod -er be.able-INF	hac-er do -INF	
		es -o, that-MASC:SG	pues. </DA>				
	'that (I) need to be able to do that'						
	Cf: A (nullpro), que (=plata), eso						
	Cb: A						
	Transition: CONTINUE						

3.5 Impersonal pronouns

Impersonal pronouns are included in the list of forward-looking centers, but they are ranked last, as per Turan (1995) and DiEugenio (1998). Turan (1995) argues that both the impersonal null pronoun and the impersonal pronoun *insan* in Turkish should either rank very low in the Cf hierarchy, or be treated as non-referential expressions and be excluded from the Cf list altogether, since they denote "members of a set whose referential identity is irrelevant" (140-1). Following Turan, DiEugenio (1998:116) proposes that quantified indefinite subjects (qis) and arbitrary plural pronouns (pro/arb) be placed at the end of the Cf list.

Three different impersonal constructions are found in Spanish: impersonal *tú*, impersonal third person plural and impersonal *se*.

3.5.1 Impersonal *tú*

The second person singular can be used impersonally (Butt and Benjamin 1994: 374).

39	a. "B"	Son, be:3PL:PRES	son be:3PL:PRES	l -os the-MASC:PL	tutoriales. </DA> exams
	'(They) are the exams'				
	Cf: examen				
	Cb: examen				
	Transition: CONTINUE				

b. Ø Tiene-s que present-ar un -o cada año.
 imp.2SG have -2SG:PRES that present -INF one-MASC:SG every year
 '(**You**) have to take one every year'
 Cf: examen (uno), cada año, **imp-tú (nullpro)**
 Cb: examen
 Transition: continue

It is interesting to note that this second person form is often used as an indirect form of reference to the speaker. In Example (39), 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. For the time being, however, we are considering it as a type of impersonal form.

3.5.2 Impersonal third person plural

Third person plural can be used impersonally when the speaker does not include him/herself or the hearer in the reference (Butt and Benjamin 1994:374). As above, impersonal third person plurals are included in the list of forward-looking centers, but they are ranked low (arbitrary plural pronouns).

40 "A" i Y que Ø se lo d -an! </DA>
 and that nullpro:3pl 3sg obj:3sg give- 3pl:pres
 'And it was given to her'
 Cf: hermana (se), programa (lo), **imp-3pl (nullpro)**
 Cb: hermana

3.5.3 Impersonal *se*

García (1975:24) identifies three impersonal *se* constructions:

- Impersonal sentence containing an inanimate nominal that is not the logical subject: *Se quemó el dulce* 'The jam was burnt' / 'Someone burnt the jam';
- Impersonal sentence containing an animate nominal preceded by *a*, in which case no subject is available for *se* to refer to: *Se fusiló a los prisioneros* 'The prisoners were shot'/'Someone shot the prisoners';
- Impersonal sentence containing no nominal: *Se vive mejor en España* 'One lives better in Spain'.

Her analysis is compatible with Turan (1995): impersonal *se* has low deixis ("the force with which the hearer is instructed to seek the referent of the pronoun"), which means that the antecedent of *se* is irrelevant (García 1975:65).

According to García, under an impersonal or person-defocussing interpretation of *se* (vs. a reflexive), *se* is used to de-focus a human entity (1975:196). For *Se quemó el dulce*, the focus is on the fact that the jam was burnt, and not on who did the burning.

Because the antecedent of *se* is irrelevant, impersonal *se* is ranked low in the Cf list.

41 Ya se te oye muy bien. </DA>
 already imp.3SG OBJ:2SG hear: 3SG very well
 'You already sound very well'
 Cf: B (te), **imp-se**
 Cb: 0

3.6 Subjects and predicates of verb to be (*ser & estar*)

The verb to be functions as a linking verb, so subjects and predicates (nominal and adjectival) of the verb to be are co-referential and only need to be listed once in the Cf list.

42 a. no, Ø 1 -a conoc-ieras, </DA>
 no nullpro:2SG OBJ-FEM:SG know -2SG:IMP:SUBJ
 'If (you) knew her'
 Cf: A (nullpro), maestra (la)
 Cb: A
 Transition: SMOOTH SHIFT

b. Ø es un cuerazo. </DA>
 nullpro:3SG be:3SG:PRES a:MASC:SG good.body
 '(She) has a good body.'
 Cf: **maestra (nullpro) = cuerazo**
 Cb: maestra
 Transition: SMOOTH SHIFT

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

43 a. "A" Ay, a poco si Ø est-ás hinchad-o. </DA>
 at little if nullpro:2SG be -2SG:PRES swollen-MASC:SG
 'If anything, (you) are swollen.'
 Cf: B (nullpro)
 Cb: 0
 Transition: NONE

b. "B" Sí, sí. </DA>
 yes yes
 'Yes, yes.'

c. No, **hinchad-o**, no, no. </DA>
 no swollen -MASC:SG no no
 'Not swollen,'
 Cf: **B**
 Cb: B
 Transition: CONTINUE

d. **Llen-o** de granitos, no, este </DA>
 full -MASC:SG of zits no eh
 'full of zits.'
 Cf: **B**, granitos
 Cb: B
 Transition: CONTINUE

4. Unresolved issues

4.1 Speech directed to a third party not in conversation

One issue to be resolved is the treatment of utterances directed to a third party not in conversation.

Following Kameyama (1998:107), the reported unit in (44d.), *que más o menos*, constitutes an embedded segment, inaccessible to centering. The reporting unit, *Æ dice*, however, is not an embedded segment and is accessible to centering. A Centering analysis shows that the speech directed to a third party must be included in the analysis since it contains the antecedent for the null pronoun.

Analysis including speech directed to a third party

44 a. *¿* Te lat -e que como quince? </DA>
OBJ:2SG beat-3SG:PRES that like fifteen
'Does fifteen (minutes) sound about right?
Cf: B (te), quince minutos
Cb: 0
Transition: NONE

b. "B" Pues no sé yo </DA>
well not know:1SG:PRES I
'Well, I don't know.'
Cf: B (yo)
Cb: B
Transition: CONTINUE

c. // Ø llev -amos como quince minutos, mamá? // </DA>
nullpro:1PL be.talking-1PL:PRES like fifteen minutes mom?
'**Mom, have we been talking for fifteen minutes?**'
Cf: nosotros, quince minutos, mamá
Cb: 0
Transition: NONE

d. Ø dice que más o menos </DA>
nullpro:3SG say:3SG:PRES that more or less
'(She) says that (we have been talking for about fifteen minutes) more or less.'
Cf: mamá (nullpro)
Cb: mamá
Transition: CONTINUE

Analysis excluding speech directed to a third party

45 a. *¿* Te lat -e que como quince? </DA>
OBJ:2SG beat-3SG:PRES that like fifteen
'Does fifteen (minutes) sound about right?
Cf: B (te), quince minutos
Cb: 0
Transition: NONE

b. "B" Pues no sé yo </DA>
 well not know: 1SG:PRES I

'Well, I don't know.'

Cf: B (yo)

Cb: B

Transition: CONTINUE

c. // Ø llev -amos como quince minutos, mamá? // </DA>
 nullpro:1PL be.talking-1PL:PRES like fifteen minutes mom?

'Mom, have we been talking for fifteen minutes?'

d. Ø dice que más o menos </DA>
 nullpro:3SG say: 3SG:PRES that more or less

'(She) says that (we have been talking for about fifteen minutes) more or less.'

Cf: mamá (nullpro)

Cb: 0

Transition: NONE

4.2 Pronouns referring to discourse segments

A second unresolved issue concerns the use of pronouns to refer to discourse segments, and how to deal with it within Centering Theory. The following example illustrates such use of pronouns. In (46c.), the demonstrative *eso* ('that') refers to the consequences of e-mail use that have been described in the two previous utterances. It is unclear how to list such "entities" as forward-looking centers.

46 a. B" Porque Ø deja-s de escrib-ir =le a
 because imp:2SG stop-2SG:PRES of write -INF=OBJ:3SG to

l -a gente </DA>
 the-FEM:SG people

'Because (you) stop writing to people'

Cf: gente (-le), gente, imp-tú (nullpro)

Cb: 0

Transition: NONE

b. y además Ø no guard-as l -as carta-s </DA>
 and also imp:2SG not keep -2SG:PRES the-FEM:PL letter-PL

'and (you) don't keep the letters either'

Cf: cartas, imp-tú (nullpro)

Cb: imp-tú (nullpro)

Transition: RETAIN

c. "A" Sí, es -o es l -o malo, </DA>
 yes that-MASC:SG be: 3SG:PRES the-MASC:SG bad-MASC:SG

'Yes, **that** is the bad thing about it'

Cf: [**dejas de escribirle a la gente y además no guardas las cartas**] (eso)

Cb: 0

Transition: NONE

Appendix A: Transcription conventions for the ISL corpus

The transcripts include a number of conventions introduced by the transcriber. These include human and non-human noises, as explained below.

CATEGORY	BRACKET
human noises	/.../ slashes
non-human noises	#...# hash marks/pound sign
silences	*...* asterisks
mispronunciations	[...] square brackets (around whole word)
	(...) parentheses (supply missing part of word or correct pronunciation of word, only inside square brackets)
transcriber comments	{...} curly braces
accent	... vertical bars/pipes
false starts	<...> angled brackets

In addition, transcriber comments include intonation, marked with one of the following at the end of the corresponding section of speech.

{period}	Falling intonation
{comma}	Slightly rising intonation, continuation of idea, and not a question
{quest}	Marked rising intonation

These comments do not reflect, or are influenced by, sentence structure. The speaker may have the intonation of a statement whether he or she is, in fact, asking a question. He or she may have the intonation of a period after a collection of words that do not, in any way, resemble a grammatically correct or complete sentence.

References

- Butt, J. & C. Benjamin (1994). *A New Reference Grammar of Modern Spanish*. Edward Arnold/Hodder Headline Group.
- Byron, D. & A. Stent (1998). A Preliminary Model of Centering in Dialog. University of Rochester CS Department. R687.
<http://citeseer.nj.nec.com/article/byron98preliminary.html>
- Cote, S. (1998) Ranking Forward-Looking Centers. In M. Walker, A. Joshi & E. Prince, Eds. *Centering Theory in Discourse*. Oxford University Press, 55-69.
- Di Eugenio, B. (1998). Centering in Italian. In M. Walker, A. Joshi & E. Prince, Eds. *Centering Theory in Discourse*. Oxford University Press, 115-137.
- Eckert, M. & M. Strube (1999). Resolving Discourse Deictic Anaphora in Dialogues. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics*, Bergen, Norway, 37-44.
<http://citeseer.nj.nec.com/miriam99resolving.html>
- García, E. (1975). *The role of theory in linguistic analysis: the Spanish pronoun system*. Amsterdam and Oxford: North-Holland.
- Grosz, B., A. Joshi and S. Weinstein (1995) Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 2 (21), 203-226.
- Kameyama, M. (1998). Intrasentential Centering: A Case Study. In M. Walker, A. Joshi & E. Prince, Eds. *Centering Theory in Discourse*. Oxford University Press, 89-112.
- Kuno, S. (1987). *Functional Syntax*. The University of Chicago Press.
- Radford, A. (1997). *Syntactic Theory and the Structure of English: A Minimalist Approach*. Cambridge University Press.
- Taboada, M. (2002a) Centering and Pronominal Reference: In Dialogue, In Spanish. *Proceedings, 6th Workshop on the Semantics and Pragmatics of Dialogue, EDILOG*. Edinburgh. September 2002: 177-184
- Taboada, M. (2002b) Foco y pronominalización en la lengua hablada: Una primera aproximación. *Documentos de Español Actual* 3-4: 173-200.
- Taboada, M. and L. Hadic Zabala (2004) What are the units of discourse structure? Segmenting discourse within Centering Theory. Manuscript. Simon Fraser University.
- Thompson, S.A. (2002). "Object complements" and conversation. Towards a realistic account. *Studies in Language*, 26 (1), 125-164.
- Walker, M. and E. Prince (1996) A bilateral approach to givenness: A hearer-status algorithm and a Centering algorithm. In Thorstein Fretheim and Jeanette Gundel (eds.) *Reference and Referent Accessibility*. Amsterdam: John Benjamins. 291-306.