

Keeping track in sign language: coreference annotation using Centering Theory

Gemma Barberà & Guillem Massó

Universitat Pompeu Fabra

____Sign Language Corpora : Linguistic Issues____

University College London – 24 July 2009

Main goals

- To establish a possible way of coreference annotation in SLs
- To describe a hierarchy of referring expressions in LSC (Catalan Sign Language)
- To start identifying the linguistic constraints that characterize coreferential expressions

Main claims

- Referring expressions can be distinguished as to the degree in which their referents are accessible at different points in a discourse
- The processing effort to process a piece of discourse varies and this is reflected in the choice of referring expressions
- Theoretical approaches predictions:
 - nonsalient or distant antecedents → anaphora coded by a lower accessibility marker (more informative)
 - salient or recently mentioned antecedent → high accessible marker (less informative)

Discourse research in OLs: Ranking of saliency

- Prince (1981): provides a taxonomy of different values of “Assumed Familiarity”
 - Gundel, Hedberg & Zacharski (1993): cognitive statuses related to the form of referring expressions in natural language discourse – “Givenness hierarchy”
 - Ariel (1988, 1990): referring expressions indicate how accessible this piece of information is at the current stage of the discourse – “Accessibility theory”
- Grosz & Sidner (1986) / Grosz, Joshi & Weinstein (1995): Centering Theory (CT)

Motivations for using Centering Theory

- CT is a processing model that relates the local utterance-by-utterance context and discourse anaphoric reference
- It is a basis to theorize about local coherence, **salience** and **choice of referring expressions**

CT: Centers & Transitions

- Centers are linguistic constructs, referents, or semantic entities that are part of the discourse model
- Each utterance has:
 - a Backward Looking Center (C_b): the most salient referent of the previous utterance that appears in the current utterance
 - a Forward Looking Center (C_f) list: a list of referents that will be projected to the following utterance
 - a Preferred Center (C_p): the most salient referent in the current utterance.
- Transitions

	$C_b(U_n) = C_b(U_{n-1})$	$C_b(U_n) \neq C_b(U_{n-1})$
$C_b(U_n) = C_p(U_n)$	<u>Continue</u>	<u>Smooth-Shift</u>
$C_b(U_n) \neq C_p(U_n)$	<u>Retain</u>	<u>Rough-Shift</u>

CT: Example

- a. Terry really goofs sometimes.
- b. Yesterday was a beautiful day and **he** was excited about trying out **his** new sailboat.
- c. **He** wanted Tony to join **him** on a sailing expedition.
- d. **He** called **him** at 6 am.
- e. **He** was sick and furious at being woken up so early.

Applications of CT

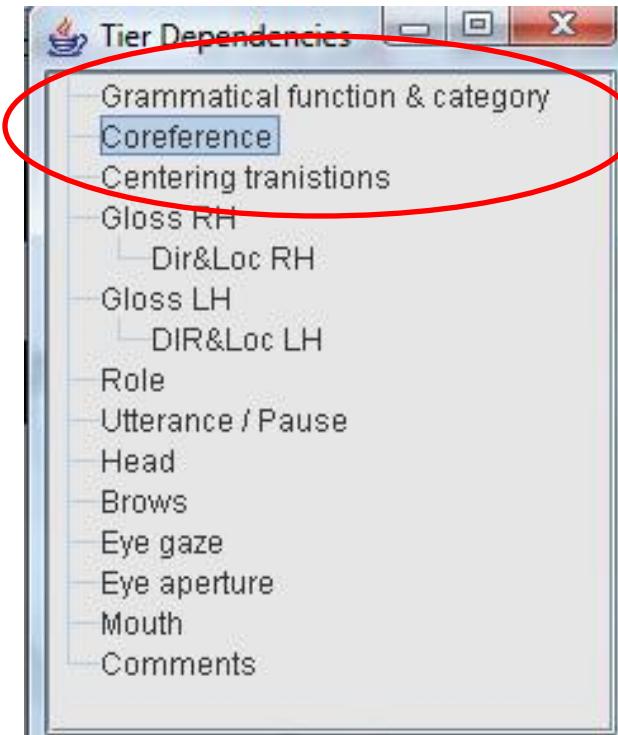
- Algorithms
 - Brennan et al. 1987
 - Walker 1989
- Applied to different OLS
 - Italian: Di Eugenio 1998
 - Japanese: Walker et al 1994
 - Turkish: Turan 1995
 - Chinese: Qinan 2008
- Anaphora resolution

Challenges with OL frameworks

- Theoretical approaches designed to study written language
- SL has features of spoken language (face-to-face interaction)
- Utterance boundaries:
 - Prosodic and interpretive cues
 - Extension of role shift
 - ... and intuition

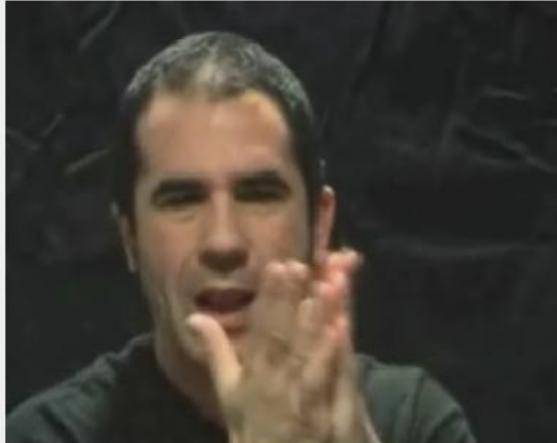
Our application

- LSC data, but useful for other SLs
- General annotation (Nonhebel et al. 2004)
- 3 linguistic tiers added



3 linguistic tiers added

File Edit Annotation Tier Type Search View Options Window Help

00:00:12.366 Selection: 00:00:11.916 - 00:00:12.136 220

Grid Text Subtitles Controls

Co-Ref

Nr	Annotation	Begin Time	End Time	Dur:
1	[H]-0 (1)	00:00:01.400	00:00:02.030	00:00:
2	[D]-0 (2)	00:00:04.696	00:00:04.926	00:00:
3	(D) (2)	00:00:05.906	00:00:06.176	00:00:
4	(D) (2)	00:00:08.966	00:00:09.256	00:00:
5	(D) (2) / (H) (1)	00:00:09.476	00:00:10.126	00:00:
6	[L]-0 (3)	00:00:12.146	00:00:12.406	00:00:
7	(D) (2) / (H) (1)	00:00:13.926	00:00:14.696	00:00:
8	(H) (1)	00:00:14.696	00:00:14.876	00:00:
9	(H) (1) / (D) (2)	00:00:15.866	00:00:16.256	00:00:
10	(H) (1)	00:00:16.346	00:00:16.636	00:00:
11	(T) 0	00:00:17.036	00:00:17.356	00:00:
12	LI (3)	00:00:17.356	00:00:17.646	00:00:
13	(L) (3)	00:00:20.686	00:00:21.666	00:00:

◀ ▶ 1◀ F◀ ▶ ▶F ▶M ▶II ▶S S' ▶ ← → ↓ ↑ Selection Mode Loop Mode

Grammatical function & category

27.000 00:00:28.000 00:00:29.000 00:00:30.000 00:00:31.000 00:00:32.000 00:00:33.000 00:00:34.000 00:00:35.000 00:00:36.000 00:00:37.000 00:00:

Grammatical fun [H] Co-Ref [D] Coreference Centering transition [C] Gloss RH [R1] Dir&Loc RH [I] DIR&Loc LH [D]

Gloss RH [R1] PU TROB 3-E D ge PU CL Man CL IL-LUSIÓ CL Manip A FANÀTI 3-EX FI PU PETONEJ PU CL Manip PET CL Mani CA FE

Gloss LH [R2] RECO M RECOR O DO D AGAFAR

Role [R1] (H) (1) (D) (2) (H) (1) (D) (2) (H) (1) (D) (2)

Annotation

1. Coreference tier:

Referring expressions realized in that utterance (list of Cfs)

2. Grammatical function / category tier:

- Subject, direct object or indirect object
- NP, CL, pronoun, null...

3. Centering transitions:

Backward Looking Center; Forward Looking Center; Preferred Center;

Type of transition

- Role tier: same index number as the coreference number for a specific referent

Referring expressions in LSC

- Full NPs & inferables
- Pronouns & index signs
- **Classifier constructions**
- Verb agreement
- Null arguments
- **(Role shift)**

Classifiers as referring expressions

- Hypothesis of categorization:
 - (i) instrumental
 - (ii) entity & limb
 - (iii) handling
- CL + topicalized NP which can occur in the same sentence or in some previous sentences
- CL are not referential by themselves. They only keep the referent **active**

CL & Role Shift

- RS indicates that the point of view holder is coreferential with a referent in the previous or matrix sentence
- CL *can* occur in role structures duplicating the referent of the point-of-view-holder (but not compulsorily so)
- When it happens this serves to assign discourse **prominence** to the referent associated with both

Idiosyncracy of Role Shift

- Role shift is usually a bridge from 3rd person to 1st person and maps anaphoric elements onto pseudo-deictic elements
- RS is a bridge from activation to saliency (to be tested empirically)

First hierarchy (to be further decomposed)

- The hierarchy of SL referring expressions is quite similar to that proposed by Ariel, Gundel at al. and Prince:
 - Full NPs
 - Entity & limb CL
 - Pronouns / verb agreement
 - Role shift
 - Null arguments

Low accessibility markers

High accessibility markers

Future work

- The ranking for the Cf in a topic-prominent language
- The role that simultaneous constructions play in the accessibility scale
- Different categorisation of CL and the precise place in the accessibility scale
- The role that the use of space plays in the accessibility scale

Thanks for your attention!!

gemma.barbera@upf.edu

guillem.masso@upf.edu

References

- AnCora-Co: Coreference Guidelines for Catalan and Spanish. Centre de llenguatge i computació de la Universitat de Barcelona. September 2008. Version 2.1
- Ariel, M. 1988. Referring and accessibility. *Journal of Linguistics*, 24, 65-87
- Ariel, M. 1990. *Accessing Noun-Phrase Antecedents*. London: Routledge.
- Brennan, S; M. Friedman & C. Pollard. 1987. A centering approach to pronouns.
In *Proceedings of the 25th Meeting of the Association for Computational Linguistics*, pages 155-162. Stanford: California
- Crasborn, O. 2007. How to recognise a sentence when you see one. *Sign Language & Linguistics* 10-2: 103-111
- Di Eugenio, B. 1998. Centering in Italian. In M. A. Walker, A. K. Joshi, and E. F. Prince (eds.), *Centering in Discourse*, 115-137. Oxford: Clarendon Press
- Grosz B. & C. L. Sidner. 1986. Attention, intentions, and the structure of discourse, *Computational Linguistics*, v.12 n.3, p.175-204
- Grosz, B. et al. 1995. Centering: A Framework for Modelling the Local Coherence of Discourse. *Computational Linguistics* 2(21), pp. 203-225
- Gundel, J., N. Hedberg & R. Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69: 274-307
- Gundel, J.K. 1988. Universals of topic-comment structure. In M. Hammond et al, eds., *Studies in Syntactic Typology*, John Benjamins, 209-239

- Kameyama, M. 1998. Intrasentential centering: A case study. In M. Walker, A. Joshi & E. Prince (eds.), *Centering Theory in Discourse*, pp. 89-112. Oxford, U.K.: Oxford
- Nespor, M. & W. Sandler. 1999. Prosody in Israeli Sign Language. *Language and Speech*. 42:2&3. 143-176
- Nonhebel, A. et al. 2004. Sign Language transcription conventions for the ECHO Project. Ms. Radboud University Nijmegen
- Poesio, M. et al. 2000. Specifying the Parameters of Centering Theory: a Corpus-Based Evaluation using Text from Application-Oriented Domains. In: *Proc. of the 38th ACL*, Hong Kong, October.
- Prince, E. 1981. Toward a taxonomy of given-new information. In Cole, P., ed. *Radical Pragmatics*. NY: Academic Press. Pp. 223-56
- Prince, E. 1992. The ZPG letter: Subjects, definiteness, and information-status. En William C. Mann y Sandra A. Thompson, *Discourse description* , 295-325. Amsterdam & Philadelphia: John Benjamins
- Turan, U. 1995. *Null vs. Overt Subjects in Turkish Discourse: A Centering Analysis*. PhD dissertation, University of Pennsylvania
- Walker, M. 1989 . Evaluating discourse processing algorithms. *Proceedings of the 27th annual meeting on Association for Computational Linguistics*, p.251-261, Vancouver, British Columbia, Canada
- Walker, M., S. Cote & M. Iida. 1994. Japanese discourse and the process of centering. In *Computational Linguistics*, v.20 June (2), p.193-232