

Deciding on units of analysis within Centering Theory

MAITE TABOADA and LORELEY HADIC ZABALA

Abstract

Many efforts in corpora annotation start with segmenting discourse into units of analysis. In this paper, we present a method for deciding on segmentation units within Centering Theory (Grosz et al. 1995). We survey the different existing methods to break down discourse into utterances and discuss the results of a comparison study among them. The contribution of our study is that it was carried out with spoken data and in two different languages (English and Spanish). Our comparison suggests that the best unit of analysis for Centering-based annotation is the finite clause. The final result is a set of guidelines for how to segment discourse for Centering analysis, which is also potentially applicable to other analyses.

Keywords: corpus annotation; discourse units; anaphora; Centering Theory; Spanish.

1. Introduction

A

(...),
(... 1986).
(...);
(... 1977);
(... 1974).
(... 1995),
(... 2004)
... A
(...)
(A ...)

A , . \ ' 30 \ ' -
 \ \ (,

1992), (1988; 1980), (1980). (1976). (1988) L2 (2001), (1992). A (1974) (), (1974: 702). A (1996), (2000) (2000: 481). (1996) (2000: 511). (1996) (2005) + (2005: 499).

⌈ (. ,) : ⌈ ⌈ -
, ⌈
⌈ (= ⌈) ;

(1997), (1997; 1997), (2004), (2004). (2004).
 (2004). A
 A 4,
)

3. Centering Theory⁶

(1995; (1998) (1986)
)

(L, 2005; L, 1998), (2004).

... forward-looking center list (L).

... preferred center (L). A backward-looking center

(L), ...

(1) ...⁷

, (1), : Harry snort. (1)

(1), ... : the

Dursleys their son, Dudley.

5.) (1) : ursleys,

udley.⁹ ... ursleys. (1)

(1) (1). (1),

...

(1), ursleys; udley, ... his

(1). ursleys, ...

(1) ursleys,

(1). (1), udley,

(1).

(1) ...

...

...

...

...

...

(2) , ɸ ():

(2) . ””

 : arry, snort

ɸ: arry – :

 : zero

 . ɸ

 . ɸ

 : ursleys, udley

ɸ: ursleys – :

 : zero

 . - \ -

 . -

 : ursleys, udley, lies, tea, member, gang, night, holidays

ɸ: ursleys – : ursleys

 : continue

 . ɸ

 ;

 : arry, udley, tea

ɸ: arry – : udley

 : rough shift

 . ɸ \ \ \ ɸ ɸ , ...

 : udley, gang, evening, park

ɸ: udley – : udley

 : continue

 . ɸ , ɸ ɸ -

 . , ɸ -

 . \ ɸ ,

 . ɸ ,

 . 1, ɸ

 . i ɸi

:
 :
 , $i-1$,
 continue,
 rough shift.
 10
 (2)
 (2) (2) (2) continue
 (2) (2) (2)
 ursleys. (2) (2),
 (2) (2) rough shift
 (2) (2) (2) continue
 : continue > retain > smooth shift > rough shift.
 2
 retain shifts, continue retain,
 (1986)
 ; (2004)
 null zero
 $i-1 =$ continue retain
 (2008)
 ;
 1.
 A , 1. 1

ursleys, (2) (his 1,
 udley). (1995).¹¹
 (1983), (1993)
 (1995) (2004),
 (1995) A () A
 A 1 (1995)

$$1 (\dots)$$

$$1, \dots, m, \dots (i-1$$

4. Variations in Centering Theory

()
 (.2004).
 . A . A
 ()
 (1995) 4.1.

4.1. Grosz, Joshi and Weinstein (1995)

(1986),
 . A
), ()
 1995, (3) (13)
 (3) (4),
 (3)
 6 A

- (5) . , ' \
- (6) .
- (7) . ,
- (8) . ' , ?
- (9),
- (9). *yeah* , A (\ 1970)
- 2006; 2008). (\)
- (9). (\)
- (9) (1994), (9)
- (9) : . \
- A: .
- : . A ,
- (10),
- ¹² *para enganchar todo* (
- (10) *porque tenemos unos por ahí para enganchar todo,*
 \ .1pl.pres .masc.pl
 .inf \ , () \ () , \ -
- (11) , \ *prefiero* ()

), - , *quedarme*

, (14),

(14)

· ' ' ' ·
·
· ' ' ' ·
· ' ' ' ·

because

(19) : . ÷

Units of analysis in Centering Theory

(1996; 1999),³

1.

$(i) \neq (i-1)$.

4

(i) = (i-1);

1987; 1998; 1999),

(2004)

(2002)

1

2. Utterances for which no topic could be identified

	1268	399 (31.47%)	1024	315 (30.76%)
	1268	404 (31.86%)	1024	318 (31.05%)
	1187	364 (30.66%)	954	278 (29.14%)
	1625	778 (47.88%)	1276	531 (41.61%)

that it breathes.

What the house does is breathe.

(1)

31.6()0()-31273.7 322.2() 31373.7 . 312.6(, ()-32273.

(24) $\text{A} > \text{A} > \text{A} > \text{A} > \text{A}$

It pleases me

(1976)

(1996)

92.89%

91.89%

407. 3

374 (91.89%). 3

422

392 (92.89%).

4. *Summary of disagreements in topic assignment*

3	/	281/338	240/332
1+2		(83.14%)	(72.29%)
:			
,		19	30
,		38	62

(406), (314, 77.34%), (281, 83.14%), (338, 83.14%), (19, 33.3%)

38.01% (46.83%)
). A
 (26), (26), (26),
 (26).
 (26) . *bueno, mandan todos los impuestos de la zona*
 .3pl.pres
 . *para que los cobremos*
 cl.3pl .2pl.subj
 . *viste, entonces bueno, funcionamos, eh,*
 .2sg.pres
 A A, (2005). A $\alpha = 0.05$
 0.01) $(p <$
 178.46) $(F_{3,10} = 118.95)^{26}$ $(F_{3,11} =$
 . A L (1953)
 $(F_{2,8} = 7.68, p < 0.05)$.

6.2. Characteristic 2: Topic = Cb

8

8. *Coincidence of topic with Cb*

	p
714 (56.31%)	570 (55.66%)
712 (56.12%)	567 (55.37%)

...

6.4. Characteristic 4: Pronoun antecedents

...

11.

11. Pronoun antecedents inside the utterance

	60 (8.01%)	68 (11.76%)
	60 (8.01%)	68 (11.76%)
	87 (11.61%)	89 (15.40%)
	13 (1.73%)	13 (2.25%)
<i>n</i>	749	578

...

($F_{3,12} = 23.52. p < 0.01$; $F_{3,12} =$

33.90, $p < 0.01$).
 $F_{2,8} = 16.08, p < 0.01$; $F_{2,8} = 36.97,$
 $p < 0.01$),

7. Discussion

(29) (29).

(5) Hay hay cosas que yo puedo hacer
 pero hay muchas cosas que no
 porque necesito instrumentos de precisión
 y yo no tengo

... (2004)

... (1987),

... (30), ...
... this, ...
... (30) ...

8. Conclusions

... () ...
... A ...
... A ...
... (1987) ...

· , ' , -

Acknowledgements

... (410-2006-1009), ...

Bionotes

A ... L ...

Notes

- 1. ...
2. ...
3. ... (2004)
4. ...
5. ... 6.2,
6. ... (1),
7. ... (2008). (2003) Harry Potter and the Order of the Phoenix.
8. ...
9. ... A ... (1986).
10. ... son Dudley. ... A ... 1983:
11. ... 195-199; ... 2000). (1995) ... 1986.

12. A \backslash : 1/2/3 - / / ; cl - ;
 sg - ; pl - ; fem - ; masc - ; pres - ;
 inf - ; subj - ; se - (,
 ,).
13. because , . , \backslash
14. (16), (I had just
 been to the bank, so I had money). so
15. ,
16. ,
17. (, -
 ,). (, -
 , (1998: 1476).
 (1979)
18. A (1981) (1995), (1979)
19. (1998)
20. A : John likes beans. As for Mary, she hates them.
 Mary. beans. , , \backslash ,
21. ,
22. A , , , - , -
 , - , -
23. (1972; , 02() 3.

References

- Ackema, P. & Neeleman, A. (1998). How to Do Things with Words. *Journal of Pragmatics* 30(1), 1-24.
- Ackema, P. & Neeleman, A. (1992). How to Do Things with Words. *Language Sciences* 14(4), 545-564.
- Ackema, P. & Neeleman, A. (1992). How to Do Things with Words. *TESOL Quarterly* 26(2), 390-395.
- Ackema, P. & Neeleman, A. (1988). How to Do Things with Words. *System* 16(2), 187-192.
- Ackema, P. & Neeleman, A. (2005). How to Do Things with Words. *Proceedings of the 43rd Meeting of the Association for Computational Linguistics (ACL-05)*, 141-148.
- Ackema, P. & Neeleman, A. (1987). How to Do Things with Words. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics (ACL-87)*, 155-162.
- Ackema, P. & Neeleman, A. (1983). *Discourse Analysis*. Cambridge, MA: MIT Press.
- Ackema, P. & Neeleman, A. (1998). How to Do Things with Words. *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL-98)*, 1475-1477.
- Ackema, P. & Neeleman, A. (1996). How to Do Things with Words. *Computational Linguistics* 22(2), 249-254.
- Ackema, P. & Neeleman, A. (1994). *Discourse, Consciousness and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Cambridge, MA: MIT Press.
- Ackema, P. & Neeleman, A. (1980). *The Pear Stories: Cognitive, Cultural and Narrative Aspects of Narrative Production*. Cambridge, MA: MIT Press.
- Ackema, P. & Neeleman, A. (2005). How to Do Things with Words. *Proceedings of Workshop in Constraints in Discourse*, 1-10.
- Ackema, P. & Neeleman, A. (2005). How to Do Things with Words. *Proceedings of Corpus Linguistics 2005*, 1-10.
- Ackema, P. & Neeleman, A. (2004). How to Do Things with Words. *Proceedings of the Third International Conference on Natural Language Generation*, 1-10.
- Ackema, P. & Neeleman, A. (2002). How to Do Things with Words. *Journal of Speech, Language, and Hearing Research* 45(6), 1232-1248.

- 2000, *Journal of Computational Linguistics* 28(1), 1–14.
- 2005 *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: John Benjamins.
- 2000, *Proceedings of the 18th International Conference on Computational Linguistics (COLING'00)*, 208–214.
- 1999, *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL-99)*, 37–44.
- 2000, *Journal of Semantics* 17, 51–89.
- 1997, *Discourse Studies: A Multidisciplinary Introduction*. London: Sage, 230–256.
- 1997, *Analysing Casual Conversation*. London: Sage.
- 2004, *Discourse Studies* 6(1), 27–52.
- 1996, *Pragmatics* 6(3), 427–454.
- 1996, *Interaction and Grammar*, 134–184.
- 1987, *Discourse Structure and Anaphora: Written and Conversational English*. London: Sage.
- 1994, *Advances in Written Text Analysis*. London: Sage, 229–249.
- 1980, *TESOL Quarterly* 14(1), 53–60.
- 2000, *GNOME Project Final Report*. London: Sage.
- 1993, *Cognitive Science* 17(3), 311–347.

1979, A \ \ : ,

- 2002 *Computational Linguistics* 28(3), 319–355.
- 2004 *Natural Language Engineering* 10(1), 25–55.
- 2005 *Design and Analysis of Experiments* (6).
- 1998 *The Function of Discourse Particles: A Study with Special Reference to Spoken Standard French*. A
- 1976 *Research in the Teaching of English* 10(1), 31–38.
- 1998 (), *Centering Theory in Discourse*. : , 327–358.
- 1997 *Computational Linguistics* 23(1), 103–139.
- 2000 *The GNOME annotation scheme manual* (). : // . . / (A).
- 2000 : A - - , *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL-2000)*. , 400–407.
- 2004 : A , *Computational Linguistics* 30(3), 309–363.
- 2001 A., *Asia Pacific Journal of Speech, Language, and Hearing* 6(1), 39– .000 - ()-3.5() / 5 1 0.97.4381 0

- 2000 .
 . *Journal of Speech, Language, and Hearing Research* 43(2),
324–339.
- 1969 *Speech Acts*. : .
- 2000 . *Language in Society* 29,
477–517.
- 2006 , A -
A . (.).

- 2006 , L . *Dis-*
course Studies 8(3), 423–459.
- 2001 A . *Com-*
putational Linguistics 27(4), 507–520.
- 2005 , A ,
(), *Anaphora Processing: Linguistic, Cognitive and Computa-*
tional Modelling. A / : , 73–95.
- 2003 A , *In-*
ternational Symposium on Reference Resolution and its Applications to
Question Answering and Summarization. , , 1–8.
- 2002 A.
2003 :

- 1970 . *Papers from the Sixth Regional Meeting of the Chicago Linguistics Society*. : \ , 567–577.
- 2002 . *Natural Language Engineering* 8(1), 69–89.