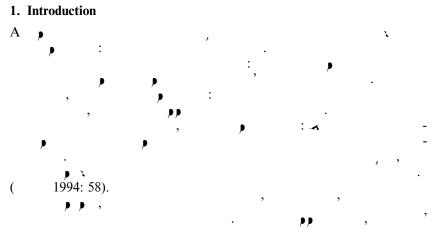
## Deciding on units of analysis within Centering Theory

MAITE TABOADA and LORELEY HADIC ZABALA

#### Abstract

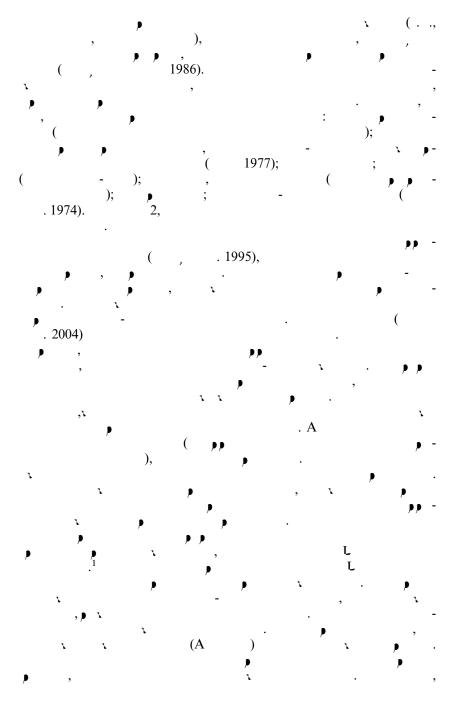
Many efforts in corpora annotation start with segmenting discourse into units of analysis. In this paper, we present a method for deciding on segmentation units within Centering Theory (Grosz et al. 1995). We survey the different existing methods to break down discourse into utterances and discuss the 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.

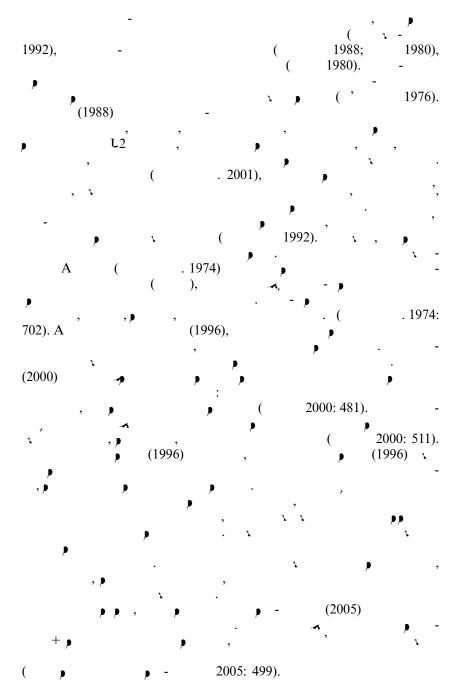


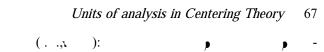
 Corpus Linguistics and Linguistic Theory 1 (2008), 63-108
 1613-7027/08/0001-0063

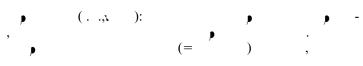
 10.1515/
 LL
 .2008.003
 ©

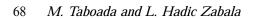


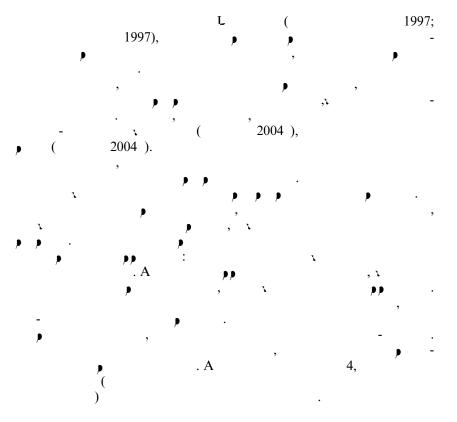
	•		,		•	-
А		r		30		,
1		1		(		



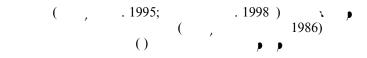


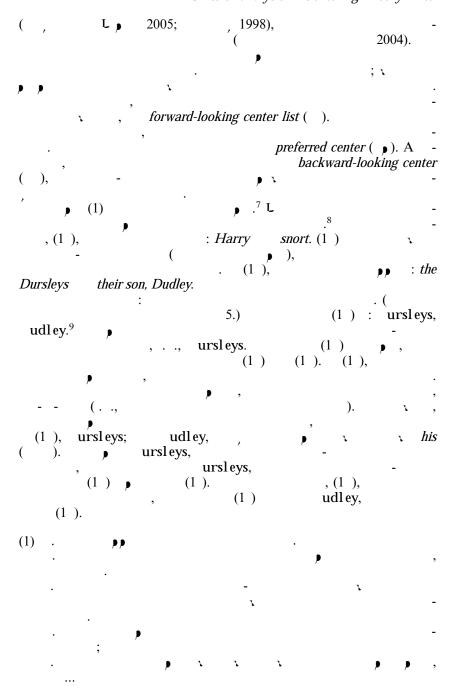






**3.** Centering Theory<sup>6</sup>



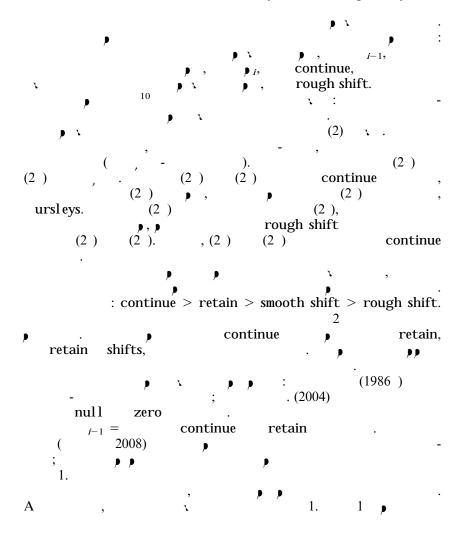


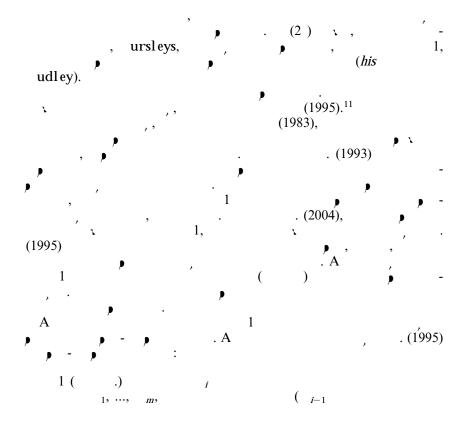
#### (2) , **)** ( ): , **þ** (2) ,, arry, snort : •: arry – : : zero • , : ursleys, udley ): ursleys – : : zero r . r : ursleys, udley, lies, tea, member, gang, night, holidays ▶: ursleys – : ursleys : continue . ۶ ; : arry, udley, tea ): arry – : udley : rough shift **)** 1 л г , ... udley, gang, evening, park udley – : udley : **)**: : continue , r 1, P ĵ

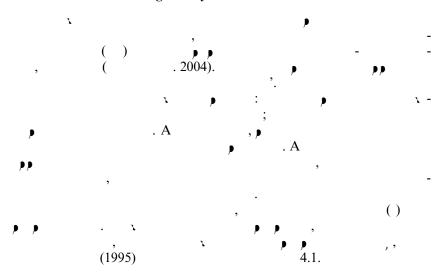
₿i

70

M. Taboada and L. Hadic Zabala

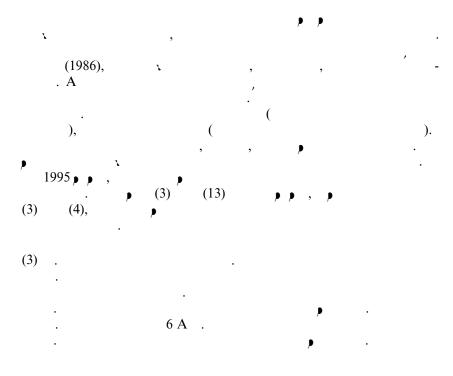


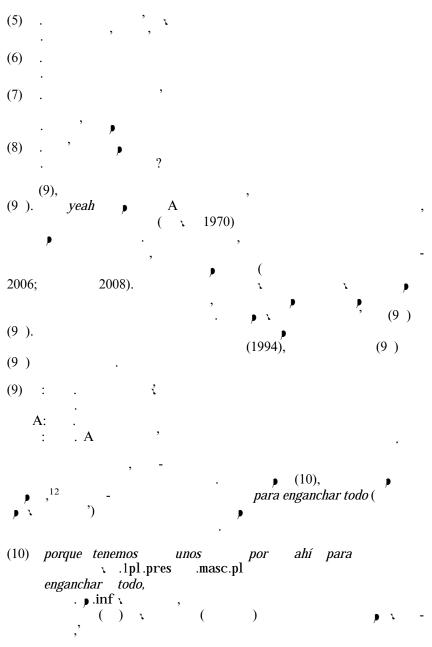




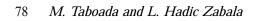
## 4. Variations in Centering Theory

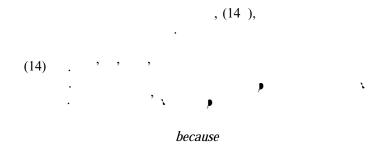
4.1. Grosz, Joshi and Weinstein (1995)



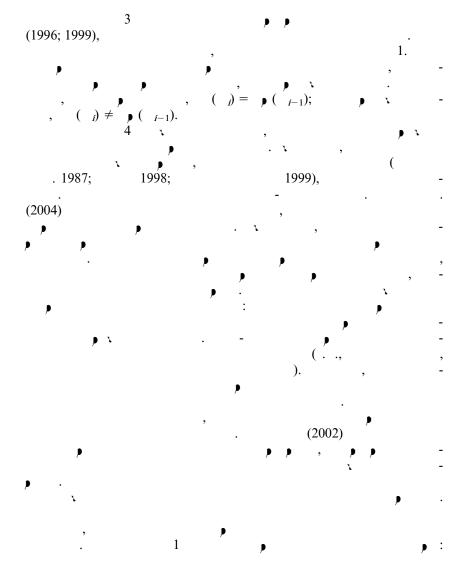


• ') - • , quedarme



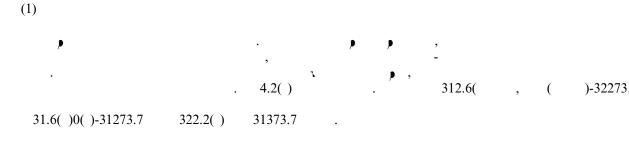


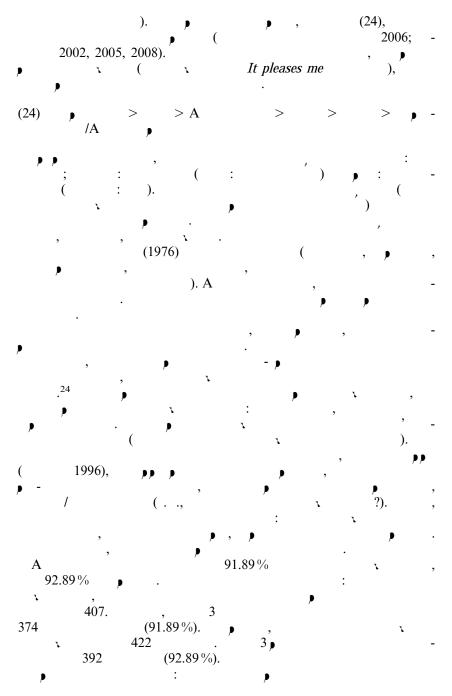
80 *M. Taboada and L. Hadic Zabala* (19) : . 、 、



		,	
	۶		۶
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%)

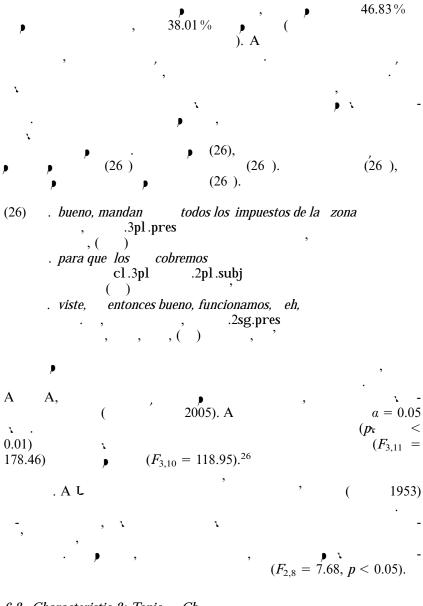
that it breathes. , , : What the house does is breathe.





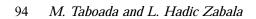
				•	
	3 1+2	/	281/338 (83.14%)	240/332 (72.29%)	
	; , P		19 38	30 62	
,	(406	), (314,	1 2 ( , 77.34 %	) 3 ).	·
	). \	83.14% 57	, 19 (3	(281 33 (33.3%)	4 . 38 - -

4. Summary of disagreements in topic assignment



8. Coincidence of topic with Cb

	۶
714 (56.31%) 712 (56.12%)	570 (55.66%) 567 (55.37%)
· · · · ·	× /





6.4. Characteristic 4: Pronoun antecedents

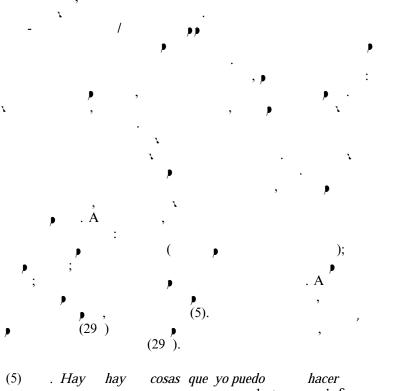


11. Pronoun antecedents inside the utterance

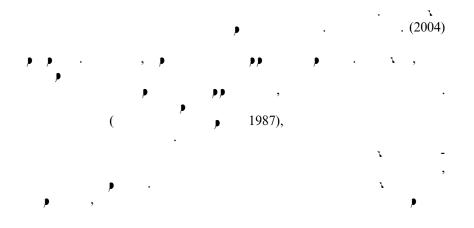
		,	
	60 (8.01%)	68 (11.76%)	
	60 (8.01%)	68 (11.76%)	
	87 (11.61%)	89 (15.40%)	
	13 (1.73%)	13 (2.25%)	
n	749	578	
•	,		
•	,		,
,	,		,
, ,	, . , ●		,
, ,	, . ♪		,
, , , , , , ,	۶,	2. <i>p</i> < 0.01; <b>)</b>	, $F_{3,12} =$

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



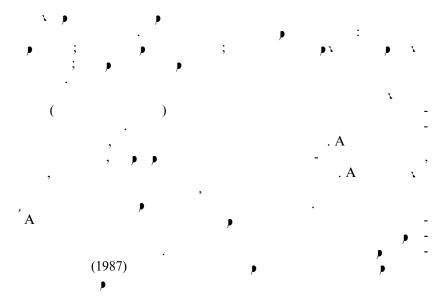


() ', . porque necesito instrumentos de precisión .1sg.pres , () , . y yo no tengo





## 8. Conclusions



· , , , , **)** –

. . . . . .

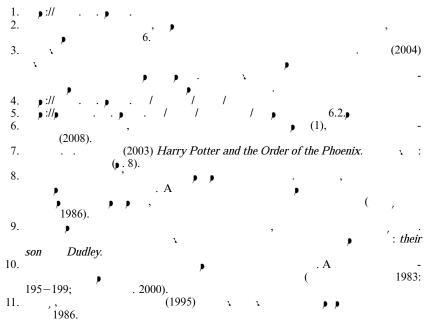


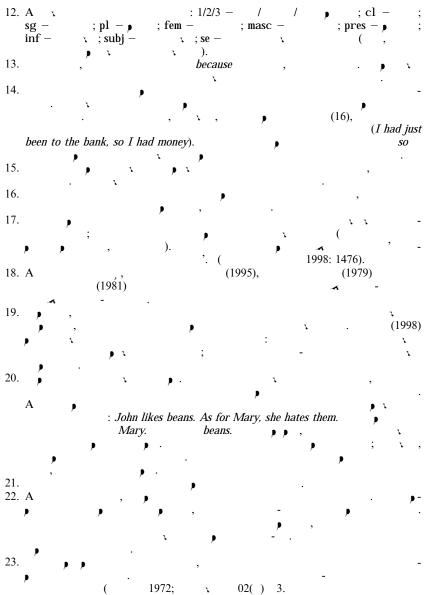


#### **Bionotes**



#### Notes





02() 3. 1972; ı

#### References

A , . 1 1962	L. How to Do Things with Words.
, 1992	545-564.
ر ۲ 1992	, A - :
, 1988	<i>System</i> 16(2), 187–192.
2005 <sup>°</sup>	L , : A - , ), Proceedings of the 43rd Meeting of the Association for Computational Linguistics. A A - , , 141–148.
1987 <sup>°</sup>	A , Proceedings of the 25th Annual Meet- ing of the Association for Computational Linguistics (ACL-87). A, A, 155–162.
1983	Discourse Analysis. :
, 1998	A , Proceedings of the 36th An- nual Meeting of the Association for Computational Linguistics (ACL-98). , , 1475–1477.
, 1996	A : . Compu- tational Linguistics 22(2), 249–254.
, 1994	Discourse, Consciousness and Time: The Flow and Displacement of Con- scious Experience in Speaking and Writing.
, 1980	( .). The Pear Stories: Cognitive, Cultural and Narrative Aspects of Narrative Production. , : A
2005 '	· · · · · · · · · · · · · · · · · · ·
2005 ,	guistics 2005. , . , Proceedings of Corpus Lin-
2004 ,	- , , Proceedings of the Third International Con- ference on Natural Language Generation. , , , . A.
2002	
	Journal of Speech, Language, and Hearing Research 45(6), 1232–1248.

, 2000	, , , , , <b>,</b> , <b>,</b> , <b>,</b>
, 2005	( .) C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Lan- guages. A / , :
2000	: A , The 18th In- ternational 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	Semantics 17, 51–89.
, , 1997 ́	as Structure and Process. Discourse Studies: A Multidisciplinary Introduc- tion. L : , 230–256.
, , 1997 ́	Analysing Casual Conversation. L :
,2004	. Discourse Studies 6(1), 27–52.
, 1996	, A. A. , 6(3), 427–454.
, 1996	A , , , , , , , , , , , , , , , , , , ,
	A ( .), Interaction and Grammar. : , 134–184.
, 1987	A. Discourse Structure and Anaphora: Written and Conversational English.
, 1994	. , , , , , , , , , , , , , , , , , , ,
, <b>)</b> 1980	. : A <b>, ,</b> , , , ,
2000	GNOME Project Final Report. : .
1993 <sup>°</sup>	-1 p, L 93 4 p. p, A: L-
, 1993	., L A. , Cognitive Science 17(3), 311–347.

, A , , , 1979 :

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	A . Research in the Teaching of English 10(1), 31–38.
1998	, A., A ,
1997	, . L . Computational Linguistics 23(1), 103–139.
2000 <sup>°</sup>	The GNOME annotation scheme manual (       ).       :         .       .       .       .       :         .       .       .       .       .       .         .       .       .       .       .       .       .         .       .       .       .       .       .       .       .         .       .       .       .       .       .       .       .       .       .         .
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. A., L
2001	Journal of Speech, Language, and Hearing 6(1), 39– . Asia Pacific .000 – ()-3.5(

) / 51 0.97.4381 (

2000	• Journ 324–339.	• nal of Speech, Langua	) ge, and Hearin	r g Research 43(2),
, 1969	Speech Acts.	:	ч.	
2000	477-517.	X	. Langi	uage in Society 29,
2006	, A - A •	۶	( 、	).

2006		: L . Dis-
		<i>course Studies</i> 8(3), 423–459.
2001	,	
2001		A p - v p . Com-
		putational Linguistics 27(4), 507–520.
2005		• . , A ,
		( .), Anaphora Processing: Linguistic, Cognitive and Computa-
		tional Modelling. A / • :
		λ (10 μαι 10 α α μαίας). Ττ (1 μαι τη μαίας) (1 τη μαίας
2002	,	
2003		A , , <i>In</i> -
		ternational Symposium on Reference Resolution and its Applications to
		<i>Question Answering and Summarization.</i> , , 1–8.
		A.
2002	,	
2002		
		2003

.

1970	•	Papers from	the	Sixth Regional Me	eet-
	ing of the Chicago Linguistics Soci 567–577.	iety.	:	1	,
2002 ′	, neering 8(1), 69–89.		. Na	tural Language Ei	ngi-