

The role of animacy in the ranking of entities in complex NPs: Evidence from Greek

This paper is concerned with discourse prominence in Complex NPs (CNPs), drawing evidence from Greek.

The results of extensive work in the interpretation of pronouns can be summarized in two main approaches. The structural approach suggests that grammatical role is directly or indirectly affecting the perceived prominence in discourse (e.g., Brennan et al 1987, Grosz et al 1995). It is observed that subjects tend to be more salient than non-subjects. The semantic approach suggests that verb semantics project focusing preferences and makes predictions with respect to thematic roles. A verb of action, for example, puts into focus the entity associated with the patient role (e.g., Stevenson et al 1994). Neither approach, however, makes any predictions regarding the salience status of entities evoked within a CNP as non-heads cannot be characterized with grammatical or semantic roles. In this study, we seek to understand the factors playing a role in making one entity more prominent than the others in CNPs and draw our evidence from Greek.

Walker & Prince (1996)'s CNP assumption and Di Eugenio (1998)'s CNP Hypothesis are two initial attempts to characterize the relative salience of entities evoked in CNPs but their hypotheses have not been evaluated empirically. Gordon et al 1999 studied possessive NPs and found that the head referent was more salient but in his data the CNPs evoked two entities of the same type (animate) so it's hard to generalize the result. Complex NPs in Greek demonstrate a variety of syntactic configurations. Head nouns can combine with genitive nouns to express possessive, (1), or other relations, (2) and they can combine with PPs, (3), and they can evoke more than two entities (4). Our dataset includes CNPs of all syntactic configurations, evoking two or more referents. Our dataset includes 484 tokens of CNPs and was constructed from a corpus of about 182,000 words compiled from the on-line publication of the Greek newspaper "Eleftherotypia". CNPs with coordinated nouns, cases of intrasentential anaphora or reference in a parenthetical sentence were also excluded from our dataset. We also excluded CNPs when none of the evoked entities was referenced in the immediately following discourse (i.e., in the next sentence). We have assumed that the entity that was referenced is more salient than the others which were not (note that we have not looked at the relationship between CNP entities and other entities in the discourse). For each entity evoked in the CNP we coded three semantic types: animate human, (AH), inanimate concrete object, (IC), (e.g., book.) and inanimate abstract objects, (IA), (e.g., freedom). Non-human animate entities were not found in our corpus. In what follows, we report the results of the CNPs of the form 'head followed by genitive'. Table 1 shows the results for all the attested combinations. The column "Ref. to H" shows times of reference to the head noun of the CNP and the column "Ref. to GN" shows times of reference to the genitive entity. Table 1 shows a strong preference for reference to the head of the CNP when the head is AH. However, the headedness effect is lost when the head is IC or IA, which suggests that AH is more prominent than IC and IA independently of headedness. Focusing on the IA-IC combination, we see a strong preference for reference to IC, which suggests that IC is more salient than IA. The analysis of the data reveals a semantic hierarchy that strongly predicts subsequent reference: AH>IC>IA. The hierarchy, of course, fails to make predictions for CNPs evoking entities of the same semantic type. In AH-AH, the head appears more prominent but the data are less clear for IC-IC and IA-IA. Further work is clearly needed to understand what determines prominence in such cases. We also looked at the type of referring expression, shown in detail in Table 2. Table 2 shows that AH referents are frequently referenced with dropped subjects or weak pronouns (the most underspecified forms of reference in Greek) whereas overall IAs and IC tend to be referenced with weak forms less frequently. We will discuss these results in the paper in more detail.

In future work, we plan to look further into whether the animacy effect persists outside the complex NP and if so how it interacts with other factors of discourse prominence. We also plan to investigate further the factors determining salience when two or more entities are of the same semantic type within or outside a CNP.

References

- Brennan, S., M. Walker-Friedman, and C. Pollard. 1987. "A centering approach to pronouns." In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pages 155–162, Stanford, California.
- Di Eugenio, Barbara. 1998. "Centering in Italian." In M. Walker, A. Joshi, and E. Prince, editors, *Centering Theory in Discourse*. Clarendon Press, Oxford, pages 115–137.
- Gordon, P. C. Hendrick R. Ledoux K. and Yang C.L. 1999. Processing of Reference and the Structure of Language: An Analysis of Complex Noun Phrases. *Language and Cognitive Processes*, 14, pages 353–379.
- Grosz B., Joshi A. and Weinstein S. (1995). Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, 21(2):203-225.
- Stevenson, R., R. Crawley, and D. Kleinman. 1994. Thematic roles, focusing and the representation of events. *Language and Cognitive Processes* 9:519–548.
- Walker, Marilyn and Ellen Prince. 1996. "A bilateral approach to givenness: A hearer-status algorithm and a centering algorithm." In T. Fretheim and J. Gundel, editors, *Reference and Referent Accessibility*. John Benjamins, Amsterdam, pages 291–306.

Examples

- | | |
|---|---|
| (1) i tsanta tis Elenis
the purse of-the Eleni-GEN
'Eleni's purse.' | (3) ta nea gia tin Ammohosto
the news for the Ammohosto
'The news about Ammohostos.' |
| (2) to plio tis katastrofis
the ship of-the disaster-GEN
'The disaster ship.' | (4) i tsanta tis thias tis Elenis
the purse of-the aunt-gen of-the Eleni-gen
'The purse of Eleni's aunt.' |

Table 1 (Referenced entities shown in blue.)

Semantic	Ref. to	Ref. to GN	Tokens total
AH-AH	27	17	44
AH-IC	32	10	42
AH-IA	8	0	8
IC-AH	13	36	49
IC-IA	12	1	13
IC-IC	18	17	35
IA-IA	20	32	52
IA-AH	15	99	114
IA-IC	5	40	45
Total	150	252	402

Table 2 (NULL S=dropped subject, WP=weak pronoun, SP=strong pronoun, Def=Definite description)

	REF TO H	TYPE OF RE	#	REF TO GN	TYPE OF RE	# OCCURENCES
AH-AH (44)	27	NULL S	13	17	NULL S	1
		WP	6		WP	10
		SP	1		SP	0
		Def	7		Def	6
AH-IC (42)	32	NULL S	15	10	NULL S	1
		WP	6		WP	1
		SP	1		SP	0
		Def	10		Def	8
AH-IA (8)	8	NULL S	3	0	NULL S	0
		WP	3		WP	0
		SP	0		SP	0
		Def	2		Def	0
IC-AH (49)	13	NULL S	1	36	NULL S	12
		WP	4		WP	11
		SP	0		SP	2
		Def	8		Def	11
IC-IA (14)	12	NULL S	2	2	NULL S	0
		WP	2		WP	1
		SP	1		SP	0
		Def	7		Def	1
IC-IC (35)	18	NULL S	3	17	NULL S	3
		WP	4		WP	3
		SP	1		SP	2
		Def	10		Def	9
IA-IA (49)	18	NULL S	5	31	NULL S	2
		WP	2		WP	9
		SP	1		SP	2
		Def	10		Def	18
IA-AH (114)	15	NULL S	7	99	NULL S	40
		WP	1		WP	25
		SP	3		SP	1
		Def	4		Def	33
IA-IC (46)	5	NULL S	2	41	NULL S	3
		WP	1		WP	7
		SP	1		SP	2
		Def	1		Def	29