

Animacy Effects on Discourse Prominence in Greek Complex NPs

Stella Tsaklidou, Eleni Miltsakaki
School of English, Aristotle University of Thessaloniki, Greece
<https://doi.org/10.36505/ExLing-2006/01/0053/000053>

Abstract

This paper is concerned with the factors determining the relative salience of entities evoked in Complex NPs. The salience of entities evoked in complex NPs cannot be predicted by current theories of salience which attribute salience to grammatical role (subjects are more salient than non-subjects) or thematic role (agent are more salient than non-agents). A plausible hypothesis might be that, in complex NPs, head nouns are more salient than non-head nouns. Based on a sizable corpus of Greek, we analyze 484 instances of complex NPs. The results of the analysis reveal a semantic hierarchy of salience which predicts the full range of data independently of headedness: Animate Human>Inanimate Concrete Object>Inanimate Abstract Object.

Introduction

The perceived prominence of entities in discourse has been analyzed extensively in a large body of the linguistic, psycholinguistic and computational literature. Discourse prominence is often correlated with the interpretation of referential expressions, especially pronouns and other referentially underspecified forms. Extensive work in this area has identified several factors responsible for making some discourse entities more prominent than others. Notably, syntactic and semantic properties of entities have repeatedly been found to correlate with discourse salience. Many researchers have observed that grammatical role is important with subjects being more salient than non-subjects (e.g., Brennan et al 1987, Grosz et al 1995). Others have observed that thematic roles are important and argued that the semantics of verbs may be responsible for bringing to focus entities instantiating specific thematic roles (e.g., Stevenson et al 2000).

The work presented here is also concerned with discourse prominence. Specifically, this paper presents a corpus-based analysis of prominence in complex NPs in Greek. Complex NPs (henceforth CNPs) are especially interesting because they evoke more than one discourse entities whose salience cannot be predicted by grammatical or thematic role. To give an example, 'John's mother' is a CNP which evokes two entities, 'John' and 'John's mother'. CNPs may evoke multiple entities which do not participate in a possession relation, for example 'garden table', 'abortion law', etc. The whole NP can be subject or object and the head referent can be an agent or a patient but

the non-head is harder to characterize in terms of grammatical and thematic role. Our aim here is to test empirically which entity (the structural head or some other entity) is perceived as more salient. The rest of this paper is organized as follows: In Section 2, we give a brief overview of prior research in CNPs. In Section 3, we present the methodology, data, results and conclusions from our corpus study of Greek CNPs.

Related work on CNPs

Prior work in entity salience in complex NPs is very limited and in most cases not tested on empirical data. Specifically, Walker and Prince (1996) proposed, but did not test, the complex NP assumption which states that in English the Cf ranking within a Complex NP is from left to right. For example, in [*Heri mother*]j knows *Queen Elizabeth*, the salience order of the entities is i>j>k. Di Eugenio (1998) proposed (but again did not test) that in possessive NPs animate possessors rank higher. If both entities in the possessive NP are animate then the head noun ranks higher. Gordon and Hendrick (1997, 1998) through a series of experiments found that the collective entity evoked by the CNP is more accessible and prominent than its component entities, noting, that this happens when the CNP is in subject position. Note that in their data, consisting of possessive NPs in English, both entities were animate.

Corpus analysis of complex NPs in Greek

The dataset for this study was constructed from a corpus of approximately 182,000 words which we collected from the on-line publication of the Greek newspaper 'Eleftherotypia'. We restricted our dataset to only those occurrences of CNP constructions which were followed by a sentence containing a reference to at least one of the entities evoked in the CNPs. We did this under the assumption that when more than two entities are evoked in the discourse, subsequent reference to one of them indicates that the referenced entity is more salient. There were no cases with subsequent reference to more than one of the entities evoked in the CNP.

The final version of our dataset consists of 484 tokens of CNPs extracted according to the following criteria: a) each CNP evokes two or more entities, and b) one of the evoked entities is referenced in the following sentence. We excluded CNPs with coordinated nouns (see Gordon & Hendrick 1997 for a related study), cases of intra-sentential anaphora (e.g., reference to one of the entities in a relative clause construction) or reference in a parenthetical sentence. In what follows we report the results of the analysis of 402 complex NPs. These are NPs evoking two entities, the head noun followed by a genitive noun. Only 82 complex NPs evoked more than two entities and we

will not discuss them further. For each entity evoked in the CNP we coded the following semantic types: animate human, (there were no animate non-human), inanimate concrete objects, (e.g., table, book, etc.), and inanimate abstract objects (e.g., freedom, honesty, etc.). Table 1 shows the results of the coding for all attested combinations. The column ‘Ref. to H’ shows how many times there was reference to the head noun of the CNP and the column ‘Ref. to GN’ shows the number of times that the referenced entity of the CNP was the entity evoked in the genitive noun.

Table 1

Semantic labels	Ref. to H	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

Table 1 shows a strong preference for reference to the head of the CNP when the head is AH, shown in (1) for AH-IC. However, the headedness effect is lost when the head is IC and the genitive noun is AH, which suggests that AH is more salient than IC. The same pattern is observed when the head is IA and the genitive noun is AH. When the head is IA but the genitive is IC we see a strong preference for reference to IC, which suggests that IC is more salient than IA. But what happens when both the head and non-head entities are of the same semantic type? In the case of AH-AH, there is a headedness effect with heads being more frequently referenced than non-heads. There is no headedness effect, though, in the IC-IC case, shown in (2), and a rather puzzling preference for reference to the non-head IA in the IA-IA cases, shown in (3). Note, however, that in (3) the referring expression is with a full noun phrase.

(1) *oi megaliteri epihirimaties, tis horas_i itan sto eleos tis dikastikis tis krisis. Orismeni apo aftous_i ine ke simera desmii tis siopis tis.*

‘The most important businessmen_i in the country_j were in the mercy of her judgment. Some of them_i are still captive to her silence’

(2) *I plioktitria eteria ihe prosthesi ki allous orofous sto plio me apotelesma na epireasti I statherotita_i tou_j. O proedros tis Egiptou dietakse ti dieksagogi epigousas erevnas gia tin eksakrivosi ton etion vithisis tou pliou_j.*

'The shipowner aimed to add some more floors on the ship with result for its_j stability_i to be affected. The President of Egypt ordered the carrying out of urgent investigation for the identification of the causes of the sinking of the ship_j'.

(3) ...sintelestike ena eglima me tous heirismous tis kivernisis ke **tin prospathia_j sigkalipsis_i. I sigalipsi_i** ine sinenohi.

A crime was committed with the manipulations of the government and the attempt_j of covering-up_i. The covering-up_i is complicity.

The analysis of the data reveals a semantic hierarchy that strongly predicts subsequent reference: **AH>IC>IA**. The results of our corpus analysis are also supported by a Centering-based study that Poesio and Nissim (2001) designed to evaluate the relative salience of entities evoked in possessive NPs in English. They compared the complex NP assumption with the left-to-right ranking with Gordon and Hendrick's finding that the head of the complex NP is more salient and concluded that actually ranking animate entities higher than ranking heads higher yields fewer Centering violations. Further studies are required to evaluate if the animacy effect that has been empirically observed in Greek and English is specific to the entities evoked in complex NPs or a more general factor for salience ranking that has been missed by the most widely accepted accounts of discourse salience.

References

- Brennan, S. Walker Friedman, M. and Pollard, C. 1987. A centering approach to pronouns. In Proceedings, 25th Annual Meeting of the Association for Computational Linguistics, pages 155--162, Stanford.
- Di Eugenio, B. 1998 Centering in Italian. In "Centering in Discourse", Prince, E. Joshi, A. and Walkers, L. editors. Oxford University Press.
- Gordon, P. C. Hendrick R. Ledoux K. and Yang C.L. 1999. University of North Carolina at Chapel Hill, USA Processing of Reference and the Structure of Language: An Analysis of Complex Noun Phrases .
- Grosz, B.J. Joshi, A.K. and Weinstein, S. 1995. Centering: A framework for modeling the local coherence of discourse. Computational Linguistics, 21(2):202--225.
- Poesio, M. and Nissim, M. 2001. Architectures and Mechanisms for Language Processing Conference, Saarbrücken, September 2001
- Stevenson, R. Knott, A. Oberlander, J. and McDonald, S. 2000. Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. Language and Cognitive Processes 15(3):225--262.
- Walker, M. and Prince, E. 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.