Cataphora, backgrounding and accessibility in discourse

Radoslava Trnavac [a,*], Maite Taboada [b]

[a] *Simon Fraser University, Department of Linguistics/Cognitive Science Program, Robert C. Brown (RCB) Building 8206, 8888 University Dr., Burnaby, B.C. V5A 1S6, Canada*
[b] *Simon Fraser University, Department of Linguistics, Robert C. Brown (RCB) Building 8206, 8888 University Dr., Burnaby, B.C. V5A 1S6, Canada*

Abstract

We examine discourse factors that are involved in the occurrence of intra/inter-sentential cataphora. On the basis of a corpus analysis of natural language, we test two cognitive theories that attempt to explain the phenomenon of cataphora: clausal backgrounding (Harris and Bates, 2002) and Accessibility Theory (Ariel, 1990). In the first part of the paper, we investigate the presence of cataphoric *he*, *she, it* and *they* in a corpus annotated with nucleus-satellite labels (Mann and Thompson, 1988), which are interpreted to be discourse correlates of the foreground-background distinction. The analysis shows that cataphora cannot be restricted to backgrounded parts of texts, and that backgrounding as an explanation for the occurrence of cataphora cannot be applied at the discourse level. In the second part of the paper, we investigate a cognitively related phenomenon to backgrounding: accessibility and its influence on cataphora. We demonstrate that in different conditions accessibility parameters such as Givenness, Distance and Unity do not show a clear influence (especially when an instance of cataphora and its antecedent are in different sentences), while Saliency and (non)Competition do play a role in the presence of cataphora (both intra- and inter-sententially).
© 2015 Elsevier B.V. All rights reserved.

1. Problem characterization

The general function of pronouns is to facilitate rapid access to the current discourse topic (Ariel, 1990; Givón, 1983; Harris and Bates, 2002). In cases of backward anaphora or cataphora, however, a pronoun is used before the referent has been introduced. Cataphora is a relatively rare phenomenon, and it has specific restrictions. Typically, cataphora occurs in a syntactically subordinate clause.[1] For instance, in Example (1a) the pronoun cannot be coreferential with the subsequent noun, but in (1b) it can, because in (1b), the cataphor occurs in a subordinate adjunct clause.

(1) a. *He** ate the cake when **the Smurf** was in the box. (From Crain (1991), cited in Harris and Bates, 2002.)
     b. When **he** was in the box, **the Smurf** ate the cake.

---

* Corresponding author. Tel.: +1 778 986 1822.
  E-mail address: rtrnavac@sfu.ca (R. Trnavac).
  [1] The term 'cataphora' in this paper is used for the pronouns *he, she, it, they*, when they linearly precede their antecedents (including false starts, repairs and right dislocations). We do not discuss here the more general phenomenon of cataphora, as described, for instance, through cataphoric devices in Gernsbacher and Jescheniak (1995).

Most of the earlier accounts have suggested that pronouns cannot precede their referents when they are the subject of the main clause (the (a) sentence in the example above), but may when the pronoun appears in a syntactically subordinate clause (the (b) sentence). On the other side, Carden (1982)[2] claims that syntactic restrictions are only symptoms of pragmatic restrictions and explores the view that an NP can appear 'to the right' of a coreferential pronoun if it answers some need for more semantic information, for instance, to re-identify a prior referent that is distant, to avoid ambiguity, to emphasize the nature of the referent, or to reintroduce the referent as topic.

Carden (1982) describes the phenomenon of backwards anaphora, or cataphora (the term that we will use throughout the paper) in the following way[3]:

$$NP_1 \ldots Pro_1 \ldots NP_2$$
$$Pro_1 \ldots NP_1$$

According to Carden (1982), in the first case, we need to test whether NP_1 or NP_2 is the antecedent for the pronoun. While in some instances the sequence Pro . . . NP is a true cataphoric relation, in others it is a re-identification of a previously mentioned referent (Bolinger, 1979), which may be considered cataphora, and which we later on will refer to as repeated mention cataphora.

In the second case proposed by Carden, the sequence represents the first appearance of the referent in the discourse. This type of cataphora is common in journalistic discourse and in fiction (van Hoek, 1997), as we see in Example (2).[4,5]

(2)   Once **it** is finished, **this new high school in Riverside, California**, will serve a mostly white, mostly middle class community. [BN]

Many cases of cataphora, such as in (1), are part of the same sentence. Other cases, however, involve reference across sentences, as can be seen in (3).

(3)   You know **she** wasn't going to claim me. **My sister** had uh, an apartment in NJ and I remember there was, she had a roommate, young girl which hindsight is 20/20 she used to say she was going to kidnap in the middle of the night, put me in bed with her. [OANC]

A possible explanation for inter-sentential cataphora, as in (3), is that the pronoun appears in a pragmatically subordinate structure, that is, a clause or sentence that is in a pragmatic relation to the clause or sentence where the antecedent appears, and where the relation is one of pragmatic subordination or backgrounding (Harris and Bates, 2002). In this paper, we attempt to establish that parallelism between intra- and inter-sentential cataphora by examining examples drawn from corpora. In addition, we apply Ariel's (1990) Accessibility Theory, and characterize instances of cataphora in terms of the discourse factors that Ariel suggests are involved in anaphoric relations. We also propose that accessibility and backgrounding are cognitively related phenomena. The goal of this paper is to answer the question posed by the examples presented so far: What discourse features trigger the occurrence of intra/inter-sentential cataphora? The paper also presents a partial refutation of the above mentioned two linguistic theories that explain licensing of cataphora—backgrounding (Harris and Bates, 2002) and Accessibility Theory (Ariel, 1990). Our approach is original in that the two theories are applied for the first time, and in a corpus-based study.

The paper proceeds as follows: In Section 2, we briefly present some well-established syntactic and discourse theories that attempt to explain the phenomenon of cataphora. In Section 3, we outline the discourse theory of coherence relations of Mann and Thompson (1988), which we use to characterize backgrounding in discourse. Section 4 describes the corpora that we used in our analysis and describes the analysis based on the theory of clausal backgrounding (Harris and Bates, 2002). In Section 5 we present the distribution of cataphora with regard to backgrounding. In Section 6 we connect the phenomenon of backgrounding and accessibility and describe the four parameters of Accessibility Theory as they are presented in Ariel (1990, 2001). Section 7 demonstrates the analysis of our corpus according to the parameters of Accessibility Theory. Finally, Section 8 presents a discussion of the analysis results, and Section 9, conclusions.

---

[2] See more on the syntactic versus pragmatic approach to cataphora in Schlenker (2005).

[3] Carden (1982) also discusses cases with a quantifier: $Pro_1 \cdots Q NP_1$. We are not concerned with such cases here.

[4] Corpus sources are provided in square brackets after the examples: OANC (Open American National Corpus), BN (Broadcast News), RST (RST Discourse Treebank). See Section 4.

[5] In the examples, the cataphoric pronoun is indicated with bold and underlining. The referent is set in bold.