

TYPOLOGICAL STUDIES IN LANGUAGE 33

---

# STUDIES IN ANAPHORA

Edited by

**BARBARA FOX**

---

JOHN BENJAMINS PUBLISHING COMPANY

# STUDIES IN ANAPHORA

Edited by



JOHN BENJAMINS PUBLISHING COMPANY  
AMSTERDAM/PHILADELPHIA



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences — Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

## Library of Congress Cataloging-in-Publication Data

Studies in anaphora / edited by Barbara Fox.

p. cm. -- (Typological studies in language, ISSN 0167-7373; v. 33)

Includes bibliographical references and index.

1. Anaphora (Linguistics) 2. Typology (Linguistics) I. Fox, Barbara A. II. Series.

P299.A5S77 1996

401'.43--dc20

96-21121

ISBN 90 272 2927 9 (hb.) / 90 272 2928 7 (pb.) (European; alk. paper)

CIP

ISBN 1-55619-641-5 (hb.) / 1-55619-642-3 (pb.) (U.S.; alk. paper)

© Copyright 1996 - John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. • P.O.Box 75577 • 1070 AN Amsterdam • The Netherlands  
John Benjamins North America • P.O.Box 27519 • Philadelphia PA 19118-0519 • USA

# Introduction

Barbara A. Fox

*University of Colorado, Boulder*

One of the early successes of work on discourse and grammar was research on reference-tracking devices (a term used in Du Bois 1980). In this research, clear correlations were found between discourse-pragmatic factors and type of reference-tracking device chosen (e.g. Givón 1983; Fox 1987; various contributions to Chafe 1980). The correlations proposed in those studies include:

- (a) topicality, as measured in part by recency of last mention, with anaphoric form, such that high topicality referents are coded by pronouns or zero, whereas lower topicality referents are coded by full noun phrases;
- (b) discourse structure with anaphoric form, such that mentions within the same discourse sequence or space are coded with pronouns or zero, while mentions not within the same sequence or space are coded with full noun phrases;
- (c) focus of attention (or cognitive state) with anaphoric form, such that what the speaker assumes the hearer is attending to is coded with a pronoun or zero, while referents that the speaker assumes the hearer is not attending to are coded with full noun phrases;
- (d) speaker attitude with anaphoric form, such that displays of highly negative or positive attitudes tend to be associated with the use of full noun phrases.

These studies, especially Givón (1983), offered some of the earliest rigorously quantitative work in discourse and grammar, setting standards for later research.

The 1980s saw an explosion of research on the topic of anaphora, and studies of anaphora have since become important to our understanding of certain kinds of cognitive processes, to our understanding of the relationships between social interaction and grammar, and to our understanding of directionality in diachronic change. The contributions to this volume represent the “next generation” of studies in anaphora — defined broadly here as those morpho-syntactic forms available to speakers for formulating reference — taking as their starting point the foundation of research done in the 1980s. These studies examine in detail, and with sophistication in method and theory, what patterns of anaphora usage can reveal to us about cognition, social interaction, and language change.

While the studies presented here explore a wide range of questions about the topic of anaphora and make use of different types of data and methods of analysis, all of the studies emerge from the following basic functional question: What are the possible relationships between form and function in natural language? A particular version of this question, namely, What are the range of functions influencing the use of given anaphoric devices, is explicitly addressed by several of the contributions to this volume (**Downing, Ford and Fox, Himmelmann, Kibrik, Klein-Andreu, Lichtenberk, Schegloff, and Tao**). More general versions of this question, examining the larger possibilities of form-function relationships, including constraints on those relationships both cross-linguistically and diachronically, are addressed in **Abraham, Frajzyngier, and Mithun**.

**Downing** explores the functions of proper nouns in English conversation, offering “territory of information” (Kamio 1994) as a factor that has been neglected by work in anaphora. **Ford and Fox, and Schegloff** suggest including social interactional factors, especially participation alignment, as functions that discourse-functional research should attend to. **Himmelmann** provides a preliminary study of the universal functions of demonstratives. **Kibrik** offers a detailed quantitative study of the discourse functions influencing the choice of pronoun over full noun phrase in Russian narratives. **Klein-Andreu** details, based on a large corpus of conversational data, the multiple uses of pronouns in modern dialects of Peninsular Spanish. **Lichtenberk** analyzes the distribution of anaphoric devices in To’aba’ita narratives. **Tao** challenges some assumptions about the functions of zero anaphora in Chinese discourse. **Frajzyngier and Mithun** are both concerned with claims of constraints on form-function relationships diachronically, most specifically with claims of

unidirectionality in language change. **Abraham** investigates typological correlations of proclitic and enclitic forms.

Perhaps because of the obviously discourse-sensitive nature of anaphora, most of the contributions to this volume investigate the functions of anaphoric devices in discourse. The discourse studies in this volume orient to the earlier research on discourse and anaphora in a variety of ways. Two of the studies, Kibrik and Ford and Fox, comment on the methodology of the earlier work. **Kibrik** provides an extensive elaboration of those quantitative methods, developing a complex weighting system for a collection of discourse functions. **Ford and Fox**, on the other hand, argue for the value of looking closely at the real-time production of a single utterance. **Schegloff**, while not directly addressing earlier functionalist work on anaphora, offers a complement piece to Ford and Fox, by examining the systematic resources which are deployed in the single instance examined in Ford and Fox. Schegloff's paper thus underscores the necessary dialectic in being accountable for the orderliness of every case and in seeing the recurrent practices, available to the analyst in aggregates of cases, which make such orderliness possible. Two more of the studies focus on the categories used in the analyses of earlier work. **Tao** discusses some issues related to the notion of "discontinuity", while **Lichtenberk** looks carefully at a locus of referential management that was given no special status in earlier work — second mention. While not using discourse data, **Langacker** proposes analyses of sentence-level pronominal uses which indicate their discourse basis. Extending the recent work of van Hoek (see, for example, van Hoek 1995), Langacker provides conceptually-based analyses for what have been thought to be syntactic constraints on pronominal anaphora (of the sort proposed by Reinhart 1983). His analyses, like van Hoek's, suggest that the constraints on pronominal use at the sentence level are not different in kind from the constraints on anaphora at the discourse level.

Much of the earlier work on anaphora was concerned with understanding the choice of pronoun versus full noun phrase in some discourse type. Some of the studies in this volume expand on our understanding of anaphora by considering other anaphoric devices. **Cumming and Ono** offer insights into the "on the fly" comprehension of noun phrases in Consumer Reports articles. **Downing** takes up the neglected class of proper nouns. **Himmelman** examines the discourse uses, cross-linguistically, of demonstratives. **Tao** focuses on zero anaphora.

The very title of this volume partakes of a distinction that two of the papers call into question, namely the distinction between anaphora and deixis. **Himmelmann** finds that anaphoric uses of deictic demonstratives appear in all of the languages he investigates and he suggests that anaphoric uses of “deictics” may be as basic as more traditional deictic uses. **Klein-Andreu**’s research suggests a fluid relationship between deixis and anaphora, in that pronoun systems can undergo changes which produce more “deictic” functions for one or more of the members of the system.

**Clancy**’s article is unique in that it attempts to provide an understanding for how children learning a language which tends to leave nominal arguments unexpressed (in this case, Korean) acquire the argument structure of verbs. Her careful examination of caretaker-child interactions reveals that Korean-speaking caretakers engage in question-asking routines which manifest the full range of argument possibilities for the verbs used. In this way we are led to see that the acquisition of argument structures is an achievement arrived at through mutual interactional work on the part of caretakers and their children.

The strength and variety of approaches and analyses represented in this volume are proof of the continued vibrancy and viability of functional approaches to syntax. We offer the volume in the hopes of furthering this rich endeavor.

## Acknowledgements

I would like to thank all of the people who participated in the Anaphora Symposium (held in Allenspark, Colorado, in May 1994) for helping to make it a stimulating event. Many thanks to Makoto Hayashi, who served as assistant in the organizing of the symposium and in the editing of this volume. The symposium was supported in part by a grant from the University of Colorado, Boulder, Graduate Committee on Arts and Humanities, and by a grant from IREX.

## References

- Chafe, Wallace (ed.). 1980. *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex.
- Du Bois, John. 1980. “Beyond Definiteness: the trace of identity in discourse.” In Chafe (ed.) 1980: 203-274.
- Fox, Barbara. 1987. *Discourse Structure and Anaphora*. Cambridge: Cambridge University Press.

- 
- Givón, Talmy (ed.). 1983. *Topic Continuity in Discourse*. Amsterdam: John Benjamins.
- Kamio, Akio. 1994. "The theory of territory of information: The case of Japanese." *Journal of Pragmatics* 21:67-100.
- Reinhart, Tanya. 1983. *Anaphora and Semantic Interpretation*. Chicago: Chicago University Press.
- van Hoek, Karen. 1995. "Conceptual reference points: A cognitive grammar account of pronominal anaphora constraints." *Language* 71:310-340.

# Contents

Introduction	vii
The Discourse-referential and Typological Motivation of Pronominal Procliticization vs. Encliticization <i>Werner Abraham</i>	1
Referential Strategies and the Co-Construction of Argument Structure in Korean Acquisition <i>Patricia M. Clancy</i>	33
Ad Hoc Hierarchy: Lexical Structures for Reference in Consumer Reports Articles <i>Susanna Cumming and Tsuyoshi Ono</i>	69
Proper Names as a Referential Option in English Conversation <i>Pamela A. Downing</i>	95
Interactional Motivations for Reference Formulation: He had. This guy had, a beautiful, thirty-two O:lds <i>Cecilia E. Ford and Barbara A. Fox</i>	145
On Sources of Demonstratives and Anaphors <i>Zygmunt Frajzyngier</i>	169
Demonstratives in Narrative Discourse: A Taxonomy of Universal Uses <i>Nikolaus P. Himmelmann</i>	205
Anaphora in Russian Narrative Prose: A Cognitive Calculative Account <i>Andrej A. Kibrik</i>	255
Anaphora, Deixis, and the Evolution of Latin <i>Ille</i> <i>Flora Klein-Andreu</i>	305
Conceptual Grouping and Pronominal Anaphora <i>Ronald W. Langacker</i>	333

---

Patterns of Anaphora in To'aba'ita Narrative Discourse <i>Frantisek Lichtenberk</i>	379
New Directions in Referentiality <i>Marianne Mithun</i>	413
Some Practices for Referring to Persons in Talk-in-Interaction: A Partial Sketch of a Systematics <i>Emanuel A. Schegloff</i>	437
Topic Discontinuity and Zero Anaphora in Chinese Discourse: Cognitive Strategies in Discourse Processing <i>Liang Tao</i>	487
Index	515

# **The Discourse-referential and Typological Motivation of Pronominal Procliticization vs. Encliticization**

Werner Abraham  
*University of Groningen*

## **1. Goals and their Motivations**

This paper seeks to show that the occurrence of *procliticization* vs. *encliticization* is motivated by the basic order of the main parts of speech in the clause. In other words, I claim that the two clitic orders are predictable order-typologically: V-marginal languages (SOV as well as VSO) will have enclitics, while SVO will develop proclitics — all to the extent that such languages exhibit this type of reduced pronominal in the first place. In the present paper I shall be concerned with pronominal clitics only in a small section of modern languages, both of the SVO and the SOV type. In order to support the claim above, two paths come to mind: First, clearly V-marginal languages, Welsh, Dutch, and German are discussed; second, a particular case will be made of Modern Cairene Arabic and Hebrew, which used to be VSO, but have developed clear SVO-traits in their modern appearances.<sup>1</sup> As SVO-languages, the Romance, mainly French and Romanian, will serve as examples in the discussion. In the end, however, I will rebuke this split between SVO and SOV/VSO as a condition for the split between PRONOMINAL PRO- and ENCLISIS, replacing it by a purely syntactic condition, i.e. that of the syntactic ‘middle field’ thereby accounting for ENCLISIS in what have always been described as SVO-languages, namely Scandinavian.

While the syntactic properties of pronominal clitics in a limited number of Indoeuropean languages have been investigated quite thoroughly (Kayne 1975 for French; Abraham and Wiegel 1993 for German and Dutch; Haegeman 1993 for Westflemish and Dutch; Cardinaletti 1992 for Italian; see further below), it seems appropriate to open the chapter with typological questions, such as: are there well-motivated positional alternatives for pronominal clitics (CL, henceforth) in SVO- vs. SOV-, or even vs. VSO-languages? And, foremost, what would be the motives for positional generalisations? These are the main questions that this article will address.

Before we tackle this let us review a few results in a checklist. This much seems to be safe ground for at least the languages that have been covered in more detail (see Kayne 1989 for French; Rizzi 1986 for Italian; Roberts 1994 for French and Italian dialects and diachronic questions in these languages; Dobrovie-Sorin 1994 for Romanian; as well as Cardinaletti and Starke 1994 and Abraham and Wiegel 1993 for Germanic and German and its main dialects, in particular).

- (1) The positions of clitics vs. their coreferential NPs and stronger pronominal forms are not identical.
- (2) Counter to Wackernagel's generalisation, there are clitic elements in first position (see Taylor 1993 for Ancient Greek or Eythorsson 1994 for a number of early Germanic languages).
- (3) The question of what can function as CL-hosts in terms of categorial properties is confusing at first sight: in SOV-languages, hosts need to be either finite verbs or AUXes (never non-finite!) or subordinating conjunctions, nothing else.
- (4) CL can, but need not, be phonological phenomena inherently (phonetically weaker forms than their non-clitic counterparts). Apart, and independently, from this, however, they are subject to syntactic (positional) restrictions. This holds for V-marginal (SOV-VSO-) languages as well as for SVO languages.
- (5) Irrespective of the SVO- or SOV-typology, CL are always in positions to the left of their fully pronominal counterparts. This is in itself a non-trivial generalization and is in need of functional explanation.
- (6) Where case distinctions play a role, such as in German, subject-CL

stand left of their clausal object partners. For the dative-accusative sequence, the CL-linearity is the reverse from that of full pronouns (and NPs). This requires a specific explanation (one that is beyond the scope of this paper; see Abraham and Wiegel 1993).

- (7) It is not clear whether predictions of any systematic type can be made with respect to the occurrence of CL in a particular language. Thus, while Serbo-Croatian and Bulgarian have ample CL-phenomena including such 'luxuries' as clitic clustering and CL-doubling, Russian, another closely related Slavic language, has no CL at all. Likewise, among the old Indo-European languages, Ancient Greek sports CL, while Latin does not.

The most prominent question to be pursued here is whether there are any regularities to be found on the basis of the Greenbergian typology or an extension thereof (and which extension or modification exactly). It is this latter question that we shall tackle in the present paper.

## 2. Pronouns and their Clitic Forms in Several Languages: Surface Typology

In the Romance SVO-languages, the finite verb appears to be an anchor point insofar as CL inevitably occur to the *left* of the verb, whereas non-CL pronominals appear in their canonical position to the right of the finite predicate (Kayne 1989; Rizzi 1986; Roberts 1989; Cardinaletti and Starke 1994, Dobrovie-Sorin 1994). To the extent that we deal with pure SVO-languages, as in the case of the Romance languages, we have a first clue as to possible generalisations: pure SVO-languages provide *proclisis*, i.e. CL-positions to the left of the finite predicate.

French (Kayne 1975):

- (6) a. \**Marie ne connaît que les* \*CL postverbally  
 Mary NEG knows except her(CL)  
 "Mary knows only her"
- b. *Marie les connaît* CL in proclisis
- c. *Marie ne connaît qu'eux* Pron in canonic governed N-position  
 Mary NEG knows except you(CL)
- d. \**Marie eux connaît* \*Pron preverbally

Of course, French *je* is a clitic since it can only occur preverbally. Postverbally, in focussed position, it has to be represented by a suppletive form (*moi*).

Italian (Cardinaletti 1994):

- (7) a. *\*Maria conosce ci* \*CL  
       Mary knows **him**(CL)  
       b. *Maria ci conosce* CL in proclisis  
       c. *Maria conosce noi* Pron in the canonic position of NPS  
       Mary knows us  
       d. *\*Maria noi conosce* \*Pron

Romanian (Dobrovie-Sorin 1994: 70):

- (8) a. *L-am rugat* CL (Pron \*il)  
       him=have(-I) asked  
       b. *\*am il/o rugat*  
       have(-I) him/her asked  
       c. *baiatul pe care l-am vazut*  
       the boy P-the **him**=have(-I) seen

The fact that the feminine pronominal CL-object occurs *enclitically* — not, however, *proclitically* as all other weak pronouns! — forces us to conclude that for cliticization phonological processes are to be separated from syntactic ones. Note, for example, the topicalized, ‘host-free’ occurrence of German *’s* as well as Dutch *’t*. See (9a-c). These are not to be taken as syntactic clitics, but as fast-speed phonetically reduced forms of pronominals. Their occurrence is not conditioned by syntactic restrictions.

- (9) a. *’s ist einmal gewesen/’s hat einst/’s atmet noch*  
       EXPL is once/ been EXPL has once/EXPL breathes still  
       “there was once/there has once been/it still breathes”  
       b. *’t is nu eenmaal zo*  
       EXPL is now once so  
       “That’s how it is”

(9a) stands for the reduced pronominal in TOP-position only as the expletive *es*, never for *es*(neuter) or *sie*(feminine). The same holds for the Dutch expletive, never for *het*(neuter), the full form, in (9b). What this shows is that there are reduced pronominals, phonetically equal to pronominal clitics, but not true syntactic clitics in any distributionally characteristic sense. What the Romanian examples below show is that conjunctions or the negative particle

also occur as a phonological host of true syntactic CL. Note that this does not alter anything with respect to the syntactic property of *proclisis*.

- (10) *Nu stie ca-l/c-o/ca-i asteapta mama*  
 not knows(-he) that=him/that=her/that=themPL expects mother  
 “He does not know that mother expects him/her/them”

If we take the Slavic languages to be V-second, but not strictly svo (i.e. with a topic position to be occupied by any clausal element, not only the subject), the question arises whether CL-phenomena pattern with the Romance regularities. See the overview below (from Dimitrova-Vulchanova and Hellan 1991):

Bulgarian

- (11) a. *Ne si li mu ja dal knigata?*  
 NEG have Q-PARTICLE **him it** given book-the?  
 “Have you not given him the book?”
- b. *Ste mu go pratja* canonic case sequence: DAT + ACC  
 FUTURE **him it** send  
 “I will send it to him”
- c. *Knigata dali mu ja dadoxa?*  
 book-the whether **him it** gave?  
 “Have they really given him the book?”
- d. *Toj kaza, ce knigata sum mu ja bil dal*  
 he said that book-the have(-I) **him it** had given  
 “He said that I had given him the book”
- e. *Toj kaza, ce na tebe knigata sum ja bil dal*  
 he said that (TO) you book-the have(-I) **it** had given  
 “He said that I had given YOU the book”

In (11a), the cl-cluster occurs according to the general sequence *li-INTERROG. PART.+AUX-CL+PRON-CL*; (11b) shows the canonic case sequence: DAT + ACC — which is important, since we shall see that in German the sequence is inverted, which will need an extra explanation. As illustrated by (11d, e), the non-clitic pronoun — the emphatic *na tebe* “to you” in the case above — occurs in a syntactic position different from CL.

## Macedonian

- (12) a. *Go vidov nego* CL-doubling (colloquial)  
**him** saw(-I) him  
 "Him I saw"
- b. *Zima ja imase pritishato Struga*  
 winter **it** had caught Struga  
 "The winter had caught the city of Struga"
- c. *Jas sum mu gi zel parite*  
 I have **him her** taken away money-the  
 "I took away the money from her"
- d. *Ja vidov Marija/zenata*  
**her** saw(-I) Mary/ woman-the  
 "I saw Mary/the woman"
- e. *Daj mu go!*  
 give **him it**  
 "Give it to him!"

## Serbocroatian

- (13) a. *Ja mu ga zelim dati*  
 I **him it** want give  
 "I want to give it to him"
- b. *Taj <pesnik> mi je <pesnik> napiso knjigu*  
 this <poet> **(to) me** has <poet> written book-a  
 "This poet wrote a book for me"
- c. *Zelim da mu ga dam*  
 want(-I) to **him it** give(-I) = 'I-want-to-him-it-give(-I)'  
 "I want to give it to him"

## Czech

- (14) a. *Ma te Jan rad?*  
 has **you** Jan in love  
 "Does Jan love you?"
- b. *Nevidel jsem te cely den*  
 NEG=seen have **you** the whole day  
 "I didn't see you all day"
- c. *Nemel jsi ho urazet*  
 NEG=should haven **him** offended  
 "You ought not have offended him"

Note that Bulgarian exhibits CL-doubling as well as CL-clustering, as in the examples below. [ $A=b$  for cliticization of  $b$  onto  $A$ , the CL-host; ‘1’, according to Indoeuropeanist tradition, for subject-nominative, ‘3’ for dative object, ‘4’ for accusative].

BULGARIAN (Dimitrova-Vulchanova & Hellan 1991):

- (15) a. *Na mene<sub>i</sub> mi<sub>i</sub> ja dadoxa knigata* Pron=CL=CL...proclisis  
to me me it gave(-they) book-the
- b. \**Mi na mene ja dadoxa knigatano* proclisis
- c. *Knigata mu ja dadoxa na Ivan* DO=CL<sub>j</sub>=CL<sub>k</sub> ... DO=DAT=ACC!  
books me it gave to Ivan
- d. \**Knigata ja mu dadoxa na Ivan* \*=ACC=DAT
- e. *Ivan mu dade na Petur knigi* NOM=DAT; clitic doubling  
Ivan him<sub>i</sub> gave to Peter<sub>i</sub> books
- f. *Ivan mi<sub>i</sub> jadade na mene<sub>i</sub> knigata* NOM=DAT=ACC; clitic doubling  
Ivan me it gave to me book-the

Both the Romance and Slavic distributions suggest, by way of the radically different and systematic positional variation of weak pronouns (CL) and strong pronouns, that there is a dependence not only of the position of the referentially identical pronouns (and full nouns), but also a positional dependence on the occurrence of CL. Does this also implicate a dependence on discourse functions, i.e. the distinction between thematic and rhematic material? I shall claim that they do.

### 3. Typological Extensions

Above, we have made a first typological generalisation, namely that pure SVO-languages, i.e. SVO without any structural space between the finite (component of the) predicate (V-fin), or the subordinating complementizer (COMPL), and the object(s) (O), show *proclisis*. We have seen that proclisis is not enacted if there is indeed such structural space between V-fin/COMPL and O, such as in the Scandinavian languages, which have *enclisis*. In order to show that there is really such a structural condition behind the phenomena, we shall look at a language type with an equally open S-O/O-S-field, i.e. with V occupying the position opposite to V-last as in the SOV-languages, which have been discussed extensively and which definitely have *enclisis*. V-initial languages in the Indo-