

Summary of Proposed Work: Anaphora in African Languages

Ken Safir, Principal Investigator - NSF grant: BCS-0303447

(This is a revised version of the short project statement provided to NSF in connection with the extension of NSF BCS-0303447, which was awarded in the spring of 2004. The NSF linguistics program funded our project with a Small Grant for Exploratory Research, SGER, which is designed as a way to provide small amounts of funds to get new projects started.)

The goal of this study is to expand our understanding of theoretically revealing empirical patterns of anaphora in the non-colonial languages of Africa and to create a web-based platform for research into these and other questions of linguistic interest.

Theoretical work inspired by the distribution of anaphora in natural language has been a major engine for innovation in linguistics for over 30 years. Although anaphora has been studied in a few African languages, none of these languages have received the kind of detailed study that has been focused on some of the major languages of Europe and Asia. Many of the most theoretically interesting aspects of anaphora that depend on subtle structural and interpretive distinctions have only been uncovered by state of the art research. For the most part, this new expertise has not yet been applied in detail to the anaphora of any African language and this project is designed to fill this gap in our knowledge. If we succeed in developing a web site that serves as a repository for thick descriptions of the anaphoric patterns of an every widening variety of African languages, we expect it will be exploited by theoreticians seeking new generalizations thereby catalyzing innovative advances in the study of anaphora.

The empirical research will be guided in part by extensions of some of the novel theoretical proposals developed in Safir (2004a) and Safir (2004b), the success of which can only be ascertained on the basis detailed empirical research of the sort envisioned. However, neither the presentation of the data nor its collection is designed to confirm or refute any particular theoretical approach. One somewhat novel aspect of his project is that it relies crucially on cooperation and collaboration of native speaker linguists as our source of primary data, relying on their linguistic training and insight both to interpret our elicitation materials and guide our attention to issues of interest. In so doing our project is intended to bring the work of linguists currently under-represented in theoretical discourse to a wide audience of linguists, both comparativists and theoreticians. Moreover, we believe that the time is right for a project like ours because the dissemination of linguistic training (both in Africa and outside of it) is creating a growing pool of potential participants to our project, and access to computers and the worldwide web, even when such access is intermittent, is enough for participants to interact with us in a timely fashion, often in spite of other complicated logistical barriers (such as dysfunctional postal systems).

Within the period this grant has covered so far, the principal investigator, Ken Safir, has developed and completed the non-language specific anaphora questionnaire (AQ) described in the initial project statement (see the AQ for discussion of how it was put together). The AQ is not designed for querying naive informants, but is rather targeted for native speaker professional linguists whose expertise is counted on for effective data-gathering. Prior to the Fourth World Congress of African Linguistics, which met at Rutgers University in June, 2003, we recruited more than 30 linguist consultants who were planning to attend the conference to work with us.

Unfortunately, 10 of the original 30 could not attend the conference and of 14 who completed a portion of the form, only four developed complete responses and participated in follow-up in the months that followed (although we remain in contact with several of the others). One native speaker linguist consultant not at the conference was recruited soon after. Uncertain of our funding after December, 2003, we stopped recruiting new consultants and concentrated on developing the case files for Yoruba, Bukusu, Urhobo, CiNsenga, and Berber as a means of refining our methodology and presentation. If new funding is forthcoming, we are prepared to engage new consultants, to develop new case files, to formulate new (additional) questionnaires, and to develop our provisional web site into a more interactive research platform.

The project as it has developed so far has proceeded in three stages of work: Recruitment, follow-up, and presentation. Recruitment has been described above and has proceeded since WOCAL 4 by a flier at conference and word of mouth. This modest outreach has brought a much greater response than we can respond to, and as a result, we have turned away potential participants with a pledge to recontact them if our funding assures both that we can remunerate them for their participation, and that we have the time and staff resources to evaluate what they send us.

In the follow-up stage, (1) the PI analyzes the questionnaire responses, looking for gaps or misunderstandings in the data developed so far, (2) the PI identifies patterns and/or phenomena that are particularly interesting as manifested in the language in question and develops further queries designed to examine said patterns/phenomena, (3) the PI or caseworkers send the new queries to our consultant(s) and a back-and-forth proceeds until the queries are concluded and returned to the PI (4) the PI and the assistant director prepare a presentation of the data for the subject language case file. The heart of the case file presentation is the completed AQ, with translations and glosses for all of the AQ queries, some based on English sentences provided, others involving sentences that the native speaker consultant is asked to produce, along with interlinear commentary by both the consultant and the project staff pertinent to understanding various details of the AQ responses. The case file also includes a grammar sketch of the subject language (often prepared by our native speaker linguist consultants), a sketch of the anaphora system in the subject language, summarizing some analytic conclusions on the basis of the AQ responses and flagging interesting analytic and theoretical issues for further investigation (prepared by the PI, our linguist consultant, or a collaboration of the two) and also a select bibliography on the subject language, usually prepared by our linguist consultant.

When the follow-up stage is complete, as it nearly is now for most of our first five cases, the PI and the assistant director turn to what needs to be done to present the language cases and the site in general to users in the most accommodating way. To facilitate crosslinguistic comparison, we plan to use the section numbers of the questionnaire as the key comparative motif along with a glossary of terms that also serve as keywords (as opposed to a simple search function, which will also be available). For example, we plan to provide a split-screen function that permits case sections to be compared side-by-side, and a keyword function will identify section numbers for every language case that the keyword matches (e.g., "bodypart reflexive" may call up a list of languages that have body part reflexives with section numbers where the matter is discussed for each language that has the phenomenon). Some keyword matches will have narrowing functions that are informed by a particular theoretical interest, such as "locality domain", which might present the reader with a set of options for domain size, then permitting a

search that will call up every language that exhibits a form sensitive to a locality domain of the relevant size (with references to the relevant section numbers for each language). These tools are not yet available, but we look forward to developing them if our funding continues.

In addition to presentation of the language cases and a guide to how the site can be used, we intend to develop the site as a platform for research on related topics, as a means of disseminating research based on our data base to the community of interested researchers (through a series of online technical reports), as a recruitment port for those who are willing to respond to the questionnaires we develop online, and as a meeting place for scholars who can come to know of each other and get in touch with each other through our site. We anticipate that the methods we employ for the study of anaphora may prove useful for the study of other issues, and so we offer our anaphora study as a paradigm for studies of other empirical issues of linguistic interest. With respect to our focus on anaphora, we are also considering improvements in our data-gathering questionnaires to facilitate less supervised work or else to explore particular areas of interest in follow-up work. As an example of the latter, we are developing a follow-up questionnaire which explores logophoricity in detail, intended for use once we discover, on the basis of our initial questionnaire, that a given language shows a morphological distinction of this kind. If our resources permit, we would also like to develop French language materials.

We anticipate that a provisional site will open for select inspection and evaluation by February 15, 2005 and that the site will be made visible to search engines sometime in March, 2005. At that time many functions of the site other than the presentation of the case files will still be in the planning stage, but these plans will also be described on the site and public discussion will be invited.