

The Role of Theory in Language Description

Edited by
William A. Foley

Mouton de Gruyter
Berlin · New York 1993

The Role of Theory in Language Description

Trends in Linguistics
Studies and Monographs 69

Editor
Werner Winter

Mouton de Gruyter
Berlin · New York

The Role of Theory
in Language Description

Edited by
William A. Foley

Mouton de Gruyter
Berlin · New York 1993

Mouton de Gruyter (formerly Mouton, The Hague)
is a Division of Walter de Gruyter & Co., Berlin.

*Dedicated to Lita Ossmundsen
Director of Research of the
Wenner-Gren Foundation for
Anthropological Research
and President from 1978 to 1986
for her pivotal role in the
furtherance of intellectual discovery
in the human sciences*

♾ Printed on acid-free paper which falls within the guidelines
of the ANSI to ensure permanence and durability.

Library of Congress Cataloging in Publication Data

The role of theory in language description / edited by William
A. Foley.
p. cm. — (Trends in linguistics. Studies and
monographs ; 69)
Includes bibliographical references and index.
ISBN 3-11-013516-7 (alk.)
I. Linguistics. I. Foley, William A. II. Series.
P125.R65 1993
410—dc20 92-46639
CIP

Die Deutsche Bibliothek — Cataloging in Publication Data

The role of theory in language description / ed. by William
A. Foley. — Berlin ; New York : Mouton de Gruyter, 1993
(Trends in linguistics : Studies and monographs ; 69)
ISBN 3-11-013516-7
NF: Foley, William A. [Hrsg.]; Trends in linguistics / Studies
and monographs

© Copyright 1993 by Walter de Gruyter & Co., D-1000 Berlin 30
All rights reserved, including those of translation into foreign languages. No part of this
book may be reproduced or transmitted in any form or by any means, electronic or
mechanical, including photocopy, recording, or any information storage and retrieval
system, without permission in writing from the publisher.
Typesetting and Printing: Arthur Collignon GmbH, Berlin.
Binding: Lüderitz & Bauer, Berlin.
Printed in Germany

Contents

Introduction <i>William A. Foley</i>	1
Noun incorporation and the nature of linguistic representation <i>Mark C. Baker</i>	13
✦ Interaction between grammar and discourse in Chicheŵa (Bantu) <i>Joan Bresnan</i>	45
✦ The elusive figures of Burmese grammar: An essay <i>A. L. Becker</i>	61
✦ A language which defies description by ordinary means <i>Andrew Pawley</i>	87
The conceptual basis of grammatical relations <i>William A. Foley</i>	131
The expanse of grammar in the “waste” of frames <i>Michael Silverstein</i>	175
Culture and conversational inference <i>John J. Gumperz</i>	193
Disjoint reference and pragmatic inference: Anaphora and switch reference in Northern Pomo <i>M. C. O'Connor</i>	215
Code, inference, placedness and ellipsis <i>Nicholas Evans</i>	243
Meaning and pragmatics of order in selected South American Indian languages <i>Doris L. Payne</i>	281

Theoretical implications of grammaticalization phenomena <i>Christian Lehmann</i>	315
Grammar and polity: The cultural and political background to Standard Malay <i>Geoffrey Benjamin</i>	341
Diglossic grammaticality in Thai <i>Anthony Diller</i>	393
Language use and language development: Review of sociolinguistic theory <i>Mohamed H. Abdulaziz</i>	421
Formalism, functionalism and the discourse of evolution <i>Jane H. Hill</i>	437
Language index	457
Subject index	461

Introduction

William A. Foley

The articles in this volume are revised versions of the papers presented in the first half of a conference entitled "The role of theory in language description", organized by the editor and sponsored by the Wenner-Gren Foundation of Anthropological Research, which was held in Ocho Rios, Jamaica, in the first week of November 1987. This conference was organized to address a few critical issues in the interface between linguistic theory and the description of natural languages, and these show up prominently in the articles in this volume, as each participant was instructed to write on an area that the organizer felt was salient to these issues.

Basically the conference concerned itself with four distinct, yet inter-connecting, questions. The first dealt with the now enormous gap between the theoretical and descriptive concerns of linguists and those of anthropologists. In the first half of this century there was a very close interaction between the two disciplines, as personified by such eminent scholars as Franz Boas and Edward Sapir, equally at home in both fields. But since about 1960, the two fields have diverged ever more widely, with the result that currently scholars in each field know very little of the recent work in the other. It was felt by the convenor that this was a situation to be deplored; indeed, that some of the most important issues in current linguistics, such as sociolinguistic variation and the role of metaphor and prototypes in linguistic meaning systems, could benefit from input from those scholars who study cultural forms and social institutions, i.e., anthropologists. From the opposing perspective, anthropologists could benefit from absorbing some of the high degree of rigor and explicitness with which linguists have staked out their bit of cultural turf — the institution of language.

The second question before the conference members related to the nature of the linguists' object of study. Much of modern theoretical linguistics has tended to abstract away from the sociocultural flux in which the game of language is played to the neurological-psychological realm, where, it is argued, the actual rules or competence for language are situated. This competence is typically highly idealized and reflects a

language-specific capacity, the “knowledge of language”, which is autonomous of both other forms of knowledge and other aspects of human social behavior. This retreat to idealized psychological structures is deemed to be necessary in order to provide an explicit and formally articulated model of language competence, removed from the vagaries and inconsistencies of human performance. Many descriptive linguists find this retreat from language as based in the human social world to be a move which renders much of the resulting theorizing irrelevant to their concerns. Whether concerned with the use of language in modern post-industrial society or the structure and use of languages spoken by small tribal groups, such linguists largely understand language as a complex, but holistic institution, one which articulates human views about the world, through representing that world or manipulating it to achieve desired ends. They, then, focus on language as an ongoing social institution in contrast to an idealized neuro-psychological capacity.

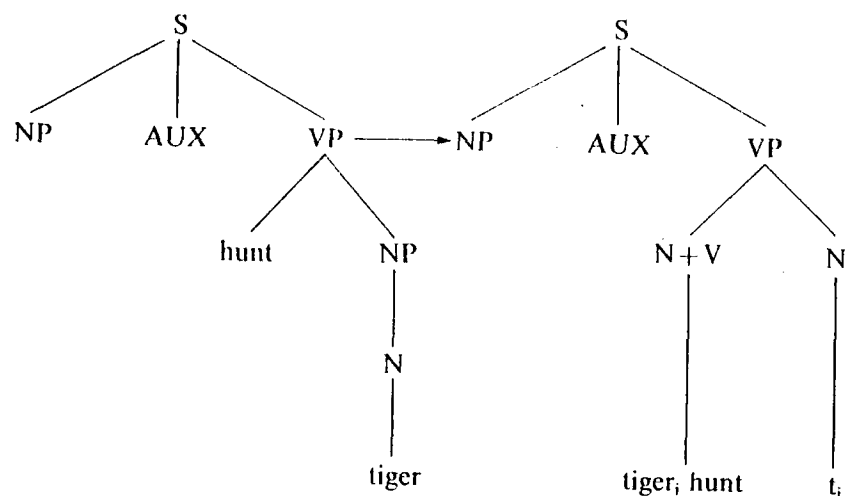
The third issue before the conference develops from the second: How to reconcile linguistic approaches which emphasize formal rigor with those which focus on pragmatic descriptions of language-based social interactions. Much of the work in theories which put a premium upon formal rigor clusters around problems in the formal representation of linguistic structures, sometimes to the point, in the eyes of more functionally oriented practitioners, of significant obfuscation of the language being described. But the central aim of such work is to specify, as formally and explicitly as possible, the constraints on the human capacity to produce and understand language structures, as part of a more general project in developing an embracing theory of mind within cognitive science. Functional approaches to language description, on the other hand, are not so focused on the capacity to produce language structures, but rather are concerned more broadly with how such structures are grounded in ongoing human social interaction. Linguists adopting this research strategy study the use and forms of language in connected discourse in actual social situations and attempt to analyze and explain the attested language structures in terms of the uses speakers make of them. Clearly these two approaches are not so much contradictory as complementary, and it was one aim of the conference to encourage practitioners of one approach to appreciate the goals of the other.

The final area of discussion concerned the data base of contemporary linguistics. The increasing concern with formal rigor in the last 25 years has gone hand in hand with a drastic narrowing of the data base of linguistic theory to languages of western Europe, most especially English.

(There have been some notable exceptions, e.g., Warlpiri, Chicheŵa, but these are remarkable for their rarity.) This work has produced a theory employing grammatical concepts, e.g., hierarchical phrase structure and grammatical relations defined on such, which are clearly based on the structure of these languages and do not generalize well to many languages of traditional or tribal communities, which are commonly of a radically different structure. Work on such languages clearly challenges many of the theoretical concepts developed on the basis of the familiar languages of Europe. The narrow focus of much of current theoretical linguistics has resulted in the view that it has little to offer the many linguists whose primary interest is in the deep and insightful description of the structure and use of the language of tribal and traditional communities. Further, the very idealization of much of current linguistic theory has rendered it largely inapplicable in meeting the very real and practical language needs of these tribal and traditional peoples and the nations in which they are found. It is crucial, if linguistics is not to become irrelevant in these contexts, that linguists meet their obligation to make their work responsive to the needs and integrity of the peoples with whom they work, peoples often socially dispossessed or disadvantaged.

In sum the goal of the conference was to try to sketch the outlines of a revitalized linguistics which would provide exciting research projects for scholars studying language as a psychological skill within cognitive science; enlighten the descriptive linguist in finding his/her way in an exotic language; offer an agenda to the sociolinguist in unravelling the institutionalization of language as a social resource; and give some practical support to the applied linguist in the areas of language planning or language teaching. It was a tall order and each of the papers had important contributions to one or more of these tasks.

The first two papers in this volume come from linguists who take a formalist perspective on language. Mark Baker's paper is within the grammatical framework of Government and Binding, the most recent model of grammar to emerge from Noam Chomsky and the MIT school. It provides an especially clear introduction to this approach. He investigates noun incorporation constructions in a variety of languages, reminiscent of English examples like *they went tiger-hunting*, where *tiger* could be viewed as an incorporated noun. Baker endeavors to explain the constraints of noun incorporation in structural terms, by arguing that incorporated nouns begin as NP arguments of the verb in a single VP. By a movement rule the nouns are shifted and adjoined to the verb to form a single complex word.



Note that this analysis suggests that the complex noun-incorporation constructions of polysynthetic languages are in their underlying forms rather like English verb phrases, and further, that aspects of the system of the syntactic categories and hierarchical phrase structure of English can be used as a proper metalinguistic vocabulary in which to couch the grammar of all languages.

Joan Bresnan's paper also would fall into the formalist camp but contrasts at many points with that of Baker. Her contribution is cast in the model of Lexical Functional Grammar, developed by herself and Ron Kaplan. Rather than taking phrasal structure and hierarchical categories of English as an apt model for universal language structure, as does Baker's Government and Binding framework, Bresnan's Lexical Functional Grammar assumes universal grammatical relations like subject, object and adjunct, as well as pragmatic discourse functions like topic and focus. Lexical Functional Grammar does not have rules like Baker's movement rule, which converts one structure into another. Rather, all sentences have three different representations and the grammar functions to forge correspondence relations between the different representations. Thus, representations involving grammatical and discursal relations are mapped onto constituent structures, but there is no derivational relationship between them. This allows a single constituent structure to fill multiple functional relations. Bresnan illustrates her framework by a detailed analysis of the pronominal prefixes to the verbs of Chicheŵa, a Bantu language of Malawi. She demonstrates significant differences between the subject and object pronominal prefixes, in their behavior and

distribution. She shows that there are systematic differences between grammatical agreement with a verb and the anaphoric relation between a morphologically bound pronoun and a discourse topic. Chicheŵa has both types of agreement relations, as shown by a variety of tests. The tests reveal that the object prefix of Chicheŵa is an incorporated pronoun used only for anaphoric agreement, while the subject prefix is functionally ambiguous between an incorporated pronoun and a marker of grammatical agreement. Finally, she argues that although the subject function is grammatically distinguishable from the topic function in Chicheŵa grammar, the subject NP is indistinguishable from the topic NP in its phrase structure properties. She concludes that it is functional ambiguity in the structural form of topic and subject constructions in Bantu that has led to the evolution of the incorporated subject pronominal into a grammatical agreement marker, and more generally, that in the architecture of universal grammar, function is not reducible of phrase-structure form.

Both Baker's and Bresnan's contributions analyze unusual phenomena in exotic languages in terms of grammatical frameworks claimed as universal models with applicability to all languages. This universalizing tendency is typical of formalist models. The papers of Becker, Pawley and Foley, in contrast, approach language description from a particularizing perspective, arguing that linguistic systems need to be understood, at least at first, deeply in their own terms, without a too hasty appeal to be meta-categories of a universalizing linguistic theory. This view is presented most strongly in the contribution from Alton "Pete" Becker. He presents a careful, particularizing analysis of the structure of the Burmese verb phrase. He argues for an especially close attention to what he calls "the discipline of the text" and cautions especially about the pitfalls of the process of translation into our own language or linguistic metalanguage, which the linguist necessarily performs to gain an understanding of the text. He points out that in our understanding of a text we always add information that the author did not intend ("exuberances") and fail to appreciate messages that the author did intend ("deficiencies"). The role of theory is to bring to the foreground both exuberances and deficiencies. In short, the process of understanding a text is a hermeneutic one, balanced precariously between the lines of the text and the lianas of the translator's language of interpretation. Becker guides the reader through this labyrinth of possible misunderstandings with his analysis of the Burmese verb phrase, which he claims contrasts with the single headed

structure of the English verb phrase in being a doubled headed one, each a pole which attracts verbal elements.

The paper by Andrew Pawley combines a particularizing and generalizing approach to understanding languages with a study of idiomatic competence in the Kalam language of Papua New Guinea. Pawley argues that much of the genius of this language is missed by a conventional "parsimonious" grammar-lexicon descriptive model, which assumes (i) that particular form-meaning pairings must be specified only once, either in the lexicon or the grammar; and (ii) that the grammar contains the productive structure-generating rules, while the lexicon is a residual list of arbitrary facts about form-meaning pairings. Kalam has a very small inventory of verbs roots (about a hundred), commonly of quite abstract meaning, and its expressive system is largely designed around this fact. Most of its conventional expressions are transparent grammatical constructions, such as verb sequences, clauses and even sentences which are well-formed but have the semantic and pragmatic force of lexemes. Pawley claims that in addition to lexemes and abstract grammatical constructions there is a third type of generative linguistic structure, which he calls "formulas". These are rather like the knowledge schemas of cognitive science and link meanings and discourse contexts to complex grammatical forms whose core lexical content (in Kalam, often a series of verbs) is partly specified. Pawley suggests that Kalam differs from better-known languages chiefly in the degree to which it exposes the shortcomings of the grammar-lexicon model as a means of describing idiomatic competence.

William Foley's paper provides a particularizing contribution on yet another language of Papua New Guinea, Yimas. Foley is at pains to demonstrate that recent claims about the universality of grammatical relations of subject and object are problematic. He considers two recent approaches to grammatical relations, one which treats them as derivative notions, defined on hierarchical phrase structure and one which takes them as primitives of grammatical theory. He then looks at data from the complex verbal morphology and cross clausal syntax of Yimas, to see if the claims of these two theories are supported. He concludes that the grammar of Yimas is organized in such a way that no unambiguous ad-hoc way of defining the grammatical relations of subject and object is possible and that neither of the two theories of grammatical relations investigated generalize well to Yimas. Rather, he concludes that the arguments of a verb must be described in semantic terms, as their participant roles, and leaves open the question as to how much this can

be done in universalist terms and how much it needs to be construed as a language particular task.

A largish number of papers at the conference investigated the role of grammar in use, i.e., the role of grammar in producing text, conversations, narratives, etc., or the role of textual functions in moulding the form of grammar. Presented with a request to contribute a paper on "How Chinook grammar shapes Chinook discourse cohesion", Silverstein responded with a direct challenge to this entire way of viewing the relation between grammar and discourse. He takes a strong stand, claiming that the notion of grammar as understood in the organizer's request is itself problematic. Rather, he argues that grammar is simply a normative abstraction created from interpretations of the complex interplay of the indexes left by "the social activity of discourse". He illustrates his points by an analysis of a Chinook text collected in 1905 by Edward Sapir, demonstrating how the "structure" of the text emerges from the complex indexical system within it. Silverstein concludes that, as discourse is a contingent, historical fact, any text produced from the realm of discourse is, as well. Texts need not develop linearly and cumulatively and, therefore, the relationship between the sentence structures of grammar and chunks of text can at best be only very indirect.

The paper by John Gumperz also emphasizes the role of inference in understanding language. In contrast to Silverstein's attention to narrative texts, Gumperz's recent work has focused on conversation. He attempts to show how both linguistic and sociocultural knowledge are used as resources by participants in a conversation to infer what is being intended at any point in time. Conversation is, then, not a sequence of sentences, speech acts or turns, but the construction of "contextualizations", through inferences accomplished cooperatively. This means that understanding is through successful negotiation in conversational events and hence effected by the economic and political forces which control access to such events. The clues which guide the process of negotiation through inference are distributed differentially within a community, and this commonly leads to misunderstanding. Gumperz concludes that the study of the conventions for "contextualizations" provides new insights into how cultural knowledge relates to grammatical knowledge and its distribution and transmission among human populations.

Catherine O'Connor's paper provides further support for Gumperz's claims. She investigates in Northern Pomo, a language of northern California, a phenomenon generally thought to be strictly syntactic — switch reference, a morphological system which indicates whether the referent

of the subject of a clause is the same or different from that of the subject of the next clause. She investigates a number of exceptions to the expected pattern, in which the different-referent suffixes are used, although the subjects are co-referential. She demonstrates that certain features of the discourse, such as backgroundness of the information presented, determine the choice, and this marked choice then guides the hearer in construing the proper inference in understanding the text. She investigates other features of anaphor in the language and argues similarly that unusual patterns are used to signal marked patterns of inference — a guide to the work of “contextualization” required by the hearer.

The paper by Nicholas Evans continues in the vein established by Gumperz and O'Connor. He proceeds to develop the notion of “placedness conditions”. These are developed to challenge the boundaries of the prevailing opposition between grammar as a “coding system” and pragmatics as an “inference system”. The “placedness conditions” are in the grammar as part of the specifications of sentences, and, again, they provide guidance in the proper constructions of inferences. Evans provides two types of situations which require “placedness conditions”: “overcoding” and “undercoding”. In “overcoding” there is a grammatical realization of the appropriateness conditions for a sentence, specifying the type of discourse in which it would be likely be found. Examples might be honorific verb forms in Japanese or the special kin-based pronominal forms in some Australian languages. “Undercoding” is best exemplified by the phenomenon of ellipsis. This leaves some of the meaning unstated, but the “placedness conditions” constrain the possible readings that can be assigned to the sentence and render unnecessary any appeal to a fully fleshed out underlying form. Evans looks at some interesting cases of normally embedded clauses being used as independent utterances (his “insubordinated clauses”) and shows how an explicit account using “placedness conditions” provides their proper interpretation.

Doris Payne's contribution continues the focus on language in action — the interplay between language structure and the actual goals of speakers in ongoing text. She investigates word order in various South American languages, specifically the arrangement between a verb and its core arguments. She concludes that word order in these languages is determined by a complex interaction of the syntactic functions of the arguments and certain pragmatic/discourse features, particularly their information status, which includes notions like focus, topic, contrast, restatement and others. Even larger discourse features like the episodic

structure of a text may be relevant. She defines these pragmatic/discourse factors in terms of the presuppositions of the speaker and hearer. These are drawn from the linguistic context; physical context of the speech event; shared background knowledge specific to the speech-act participants; shared cultural knowledge and others. Word-order variations in these languages, then, can be viewed, along the lines of Gumperz's and O'Connor's contributions, as signposts in guiding the hearer to proper inferences in understanding the text. Thus, the interaction of factors which determine the word order in any given sentence in a particular language can be complex, and the wide differences among the languages she investigates are due to the differential effects of the syntactic factors and various pragmatic/discourse factors in each language.

Christian Lehmann's paper parallels the last five in emphasizing the understanding of language as a system in use, but rather than focusing on it as a resource for producing and comprehending text, he is concerned with it as a dynamic activity which is constantly in flux. At any given moment some grammatical patterns are fading, while others are on the rise. This entails understanding grammar not as a system of fixed categories, but as a fluid mapping device between language functions and their realizations. He supports his view with a discussion of the process of grammaticalization, drawing examples from Mayan languages. Grammaticalization is a process whereby independent lexical items come to function as grammatical morphemes thus creating structure. Lehmann claims that the reality of grammaticalization demonstrates that the difference between lexical and grammatical structure, between what is conveyed and the way it is structured, must be understood as a fluid one and needs to be couched in terms of continua rather than fixed categories. A language description will achieve its goal of rendering the language system understandable, i. e., of revealing the way the language works, to the extent that it systematizes observable grammatical variation along such continua.

The three papers by Benjamin, Diller and Abdulaziz deal with language in relationship to surrounding social institutions. Geoffrey Benjamin's contribution emphasizes that the human social institution called language has many features in common with other such institutions, particularly its property as a ready-made system for the encoding of values viewed as appropriate by the larger society and polity. Benjamin divides cultural regimes according to what he calls their “modes of orientation”, with a basic dichotomy between relatively condensed, insider-oriented and event-highlighted modes (self-oriented) versus more articulated, outsider-ori-

ented and participant-highlighted modes (other-oriented). The former is less directive to the interlocutor than the latter. Languages and varieties of language can be seen as being more expressive of one mode or orientation than the other. Benjamin investigates the dialect continuum of the human institution called the Malay language, especially concentrating on the differences between rural dialects spoken by relatively traditional communities and the standardized national language, codified by (and expected to be used in) bodies of the national polity. By a detailed study of the meanings and use of the verbal affixes, he indicates striking differences in the two varieties of the language, correlated to the differences in the two modes of orientation. Benjamin notes that increased prominence of the other-oriented mode in the standardized national language can be understood in the light of its role as the articulating vehicle for the ideology of the dominating and nation-forming polity.

Anthony Diller's paper looks at the codification over the last 200 years of another language of South East Asia, Thai. Diller presents an autobiographical account of his learning of the language and the various understandings he has had of it over some three decades of experience. In his first two decades two different views of the language had emerged. One, as having an almost English-like syntactic structure, with fixed order, strict grammatical relations and subcategorization constraints, and familiar word classes; and another, almost non-configurational picture, with pragmatically controlled word order, a blurring of the distinction between lexicon and syntax, and fuzzy categories. In his third decade he has come to see both of these views as correct, but with the differences attributable to different varieties of the language. Colloquial Thai, largely spoken in the rural areas, does resemble the rather loose structure of standard non-configurational languages like Warlpiri or Nunggubuyu, but formal Thai, spoken by elites in urban areas and functioning as the national language, is much more heavily and rigidly syntactic in its organization. And further, the higher the register of formal Thai one speaks, the more rigidly syntactically structured will its sentences be, and, indeed, will approach a word-for-word translation of the corresponding English utterance. Diller points out that the source of this may be the Europeanizing influence of the kings of Thailand in the nineteenth century, who actually prescribed grammatical rules along the models of European languages. The parallel to Benjamin's discussion of "modes of orientation" and their role in the standardization of languages for modern polities is inescapable.

The final paper concerned with the role of language in the modern nation-state is that of Mohammed Abdulaziz. He is concerned with the usefulness of modern sociolinguistics in aiding modern multilingual African nations striving for development. Abdulaziz surveys broadly the great diversity of sociolinguistic situations in African countries and poses many questions as to how sociolinguistic theory can be marshalled to meet the needs of these countries in language planning and development. He particularly focuses on the problems of very multilingual nations, such as Kenya and Tanzania (with over 100 local languages) and their attempts to promote a single, unifying national language, in these two cases, Kiswahili. What is the role of the smaller, local languages? Should they be promoted, and at what cost in both economic and political terms? Abdulaziz points out that the desire of linguists to preserve a rich reservoir of linguistic diversity in these countries may conflict with the aspirations of their economic and political leaders, who want to promote a strong national language, both for the purposes of political unity and for an easier access to the benefits of economic development, which requires competence in national and international languages.

The last paper in the volume by Jane Hill looks at the old question of the evolution of language and the differing perspectives formally and functionally oriented linguists take on it. Hill starts off from recent work in biology, questioning the strict adaptionist paradigm as a sufficient explanatory basis for evolution. The adaptionist paradigm is held to be a discursive project which yields narratives ("just-so stories"), rather than testable hypotheses. This is the "Panglossian paradigm". Things are as they are because this is a functional adaptation to the conditions ("the best of all possible worlds"). The narratives of adaptionism cohere by explaining the resolution of the imbalances in the system to a Panglossian optimum. Hill then discusses recent proposals about language evolution from both the formalist and functionalist camps and demonstrates that both accounts are best characterized as discursive narratives ("just-so stories"), suffering from the internal fallacies of Panglossian optimism. She offers the possibility that some of the structure of language may more closely resemble *Baupläne*, simple autonomous formal constraints, rather than adaptionist traits. Finally, Hill's work provides the tantalizing view that not only are languages social institutions which encode appropriate ideological content, but that the same may possibly be true of our theories of language as well. This is a rather heady thought on which to conclude a volume entitled *The Role of Theory in Language Description*; one rather has visions of William James' turtles resting on the backs of other turtles, on into infinity.

Noun incorporation and the nature of linguistic representation*

Mark C. Baker

1. Introduction

Along with their other virtues, children are amazing linguists. They have (in a manner of speaking) solved the great problem of linguistic description: How should language be described given that human languages vary greatly? This is seen by the fact that any human child can learn any human language as long as it is raised in the context of that language. Thus, there is at least one important bound on the diversity of human language: all languages must be within the limits of human cognitive capacity, when the human starts learning "from scratch". Generative linguistics is largely devoted to exploring the implications of this fact.

A central linguistic controversy concerns whether linguistic representation should be abstract or concrete. As theories about how native speakers actually represent their language mentally, both approaches can account for children's facility in learning language without explicit training. If knowledge of language is concrete, then they would be able to acquire it by direct analogy and generalization from the utterances that they hear around them. This could be called the "what you see is what you get" approach to language acquisition. On the other hand, if knowledge of language is largely abstract, then there could be universal principles underlying the language in the children's minds from the beginning. Then, they need only establish the link between these innate structures and the characteristic vocabulary and patterns of their languages in order to have complete knowledge of those languages. This would be an "I knew it all along" approach to language acquisition. Either way, language acquisition can potentially be accounted for.

This said, the big question is: Which is the native speaker's linguistic description — abstract or concrete? Now, there are serious foundational questions about what the adjectives in this question mean (see, for example, Hill this volume). This notwithstanding, I will use the phenomenon of Noun Incorporation to argue that (one aspect of) a speaker's

representation must in fact be “abstract”. Descriptively, Noun Incorporation (NI) structures are those in which a nominal root is morphologically combined with a predicate to make a single, complex word. (1) is a simple example.

- (1) *Wa-hi- 'sereht- anvhsko* (Mohawk, Mithun 1984)
 past-he/me- car- steal
 ‘He stole my car.’

NI is a characteristic of many polysynthetic languages (Sapir 1911), that is, languages in which concepts are expressed largely by building very long words out of morphemes, rather than by combining words to make phrases and sentences. Many of the native languages of the Americas are strongly polysynthetic in this sense, in contrast with the more “isolating” languages brought by the Europeans.

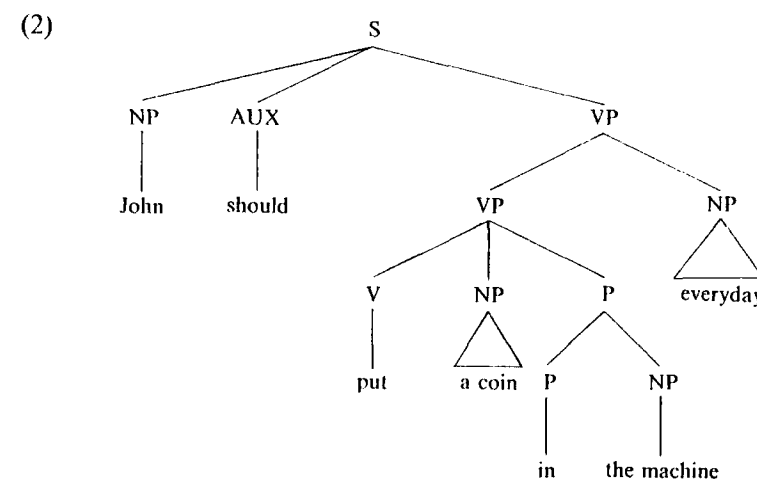
Superficially, these two linguistic systems seem almost incomparable. Nevertheless, I will show that there are important generalizations that hold of both systems – generalizations that can only be understood properly if both are in fact the same at an abstract level of linguistic representation. In this, I seek to demonstrate what I mean by the term “abstract level of representation” before I define it. The syntactic analysis of NI will be couched in the terms of Chomsky’s (1981) Government and Binding theory (GB). As the argument unfolds, I will be particularly explicit about the idealizations that are necessary and where they enter into the reasoning. Then, after it is clear what kinds of generalizations need to be captured, I return to a more careful discussion of what it means for a linguistic representation to be abstract, distinguishing between different senses of that notion.

2. The basic generalization

First, I would like to establish the following generalization (called “G”).

G: A noun can be incorporated into another category in the system of a polysynthetic language only if a noun phrase headed by that noun would be the sister¹ of the category in the phrase structure system of an isolating language.

If correct, G implies that polysynthetic and isolating languages are not incomparable systems. The phrase-structure tree in (2) illustrates the structural relationships that are widely accepted for the more isolating languages like English; this can be used as a reference point in evaluating G.



Here I abstract away both from minor differences among theoretical analyses and from variation among isolating languages (notably that found in how the words are ordered within the phrases). These are for the most part irrelevant to the interpretation of generalization G.

Already, some important idealizations are needed to clarify the intended meaning of statement G. First, it assumes that one can recognize what a noun is in the language. For example, by “noun” I mean to include full nominal roots but not so-called “incorporated pronouns” (also called subject and object agreement morphemes; cf. Bresnan 1988), which observationally have a somewhat different distribution. This type of idealization is simply an effort to narrow in on the natural class of things about which something meaningful can be said. Following Sapir (1911), I also exclude from consideration incorporated body-part terms. This is primarily a pragmatic decision; incorporated body part terms are very common, but their role in the sentence as a whole is not always clear. Finally, G gives only a necessary condition on NI, and the phenomenon is somewhat more restricted than this in some languages. These additional restrictions will also be put aside in this study, based on the intuition (or

the hope) that they can be attributed to independent, language-particular factors.

Consider now the following statement about when Noun Incorporation can occur, taken from Mithun (1984: 875). This statement sums up the results of her extensive crosslinguistic survey of NI phenomena, and is largely supported by my own investigations.

V-internally, IN's [Incorporated Nouns-MCB] bear a limited number of possible semantic relationships to their host V's... If a language incorporates N's of only one semantic case, they will be patients of transitive V's — whether the language is basically of the ergative, accusative, or agent/patient type. Turkish and the Oceanic languages illustrate this. If a language incorporates only two types of arguments, they will be patients of transitive and intransitive V's — again, regardless of the basic case structure of the language. The majority of incorporating languages follow this pattern. Many languages additionally incorporate instruments and/or locations, such as Nahuatl, ... Takelma, ... and Sora.

Putting aside for now the statement about instruments and locations, we notice that transitive verbs can express their patient argument as an incorporated noun root but not their agent argument. The following examples from Mohawk (Iroquoian) and Southern Tiwa (Kiowa-Tanoan) illustrate this.

- Mohawk (based on Postal 1962)
- (3) a. *Yao-wir-a?-a ye-nuhwe?-s ne ka-nuhs-a?* NO NI
PRE-baby-SUF she/it-like-ASP DET PRE-house-SUF
'The baby likes the house.'
- b. *Yao-wir-a?-a ye-nuhs-nuhwe?-s* OBJECT NI
PRE-baby-SUF she/it-house-like-ASP
'The baby likes the house.'
- c. **Ye-wir-nuhwe?-s ne ka-nuhs-a?* SUBJECT NI
she/it-baby-like-ASP DET PRE-house-SUF
'The baby likes the house.'

- Southern Tiwa (Allen — Gardiner — Frantz 1984)
- (4) a. *Hliawra-de Ø-k'ar-hi yede* NO NI
woman-SUF A:A-eat-FUT that
'The woman will eat that.'
- b. *Ti-seuan-mū-ban* OBJECT NI
1sS:A-man-see-PAST
'I saw the man.'

- c. **Ø-hliawra-k'ar-hi yede* SUBJECT NI
A:A-woman-eat-FUT that
'The woman will eat that' (OK as 'She will eat that woman!')

If we compare this situation to that which prevails in languages like English, we notice a two-step correspondence. First, two-argument verbs have their agents as subject and their patients (or themes) as object, not vice versa.² Second, the object noun phrase is the phrase structure sister of the verb, while the subject is not (see (2)). Thus, this basic observation about NI is perfectly consistent with generalization G.

In fact, G's structurally based formulation is perhaps more accurate than Mithun's semantic characterization of incorporated nouns, if the latter is interpreted as a theoretical claim as well as a descriptive statement. G accounts immediately for examples like the following, from the Mohawk text in Hewitt (1903: 270).

- (5) a. *Hakare' nen' ia'-e'-hent-ara'ne' ka'-hent-owane'*
after now DIR-3FS:field-reached PRE-field-large
'Then, after a while she reached a field that was large.'
- b. *O' nakarontote'nene' karonto'ne dji*
what part-PRE-tree-SUF PRE-tree-SUF where
teieita-'hia-tha'
IMP-stream-cross-INSTR
'What kind of tree is used to cross the stream there?'

In these examples, it is not the patient/theme argument — the argument in motion — that appears within the verb, but rather what one might call a "path" argument. Thus, these examples do not fit Mithun's primary generalization. They do obey G, however, since locative NPs of precisely this type can be objects of the verb in isolating languages, as the English glosses in (5) illustrate.

According to Mithun's generalization, certain other types of Ns do not incorporate into the verb. For example, nominal heads of NPs used as adjuncts never appear inside the verb in polysynthetic languages. I know of no examples which would have literal and free glosses like those in (6).

- (6) **???*
Baby AGR-day-cough-ASP [every-]
'The baby coughs every day.'

Similarly, many locative expressions cannot be incorporated into the verb.

- (7) *???
 She AGR-bush-threw-ASP him into-
 'She threw him into the bushes.'

These gaps in the range of what can be expressed polysynthetically also follow G, since neither adjuncts nor locative expressions are characteristically sisters of the verb in isolating languages. Rather, adjuncts are outside the constituent containing the verb and its complements; locatives are inside an additional constituent, namely one headed by an adposition (cf. (2)).

More idealizations worth discussing arise in relation to these statements. Thus, in (6) it is important to distinguish between adjunct incorporation and adverb incorporation. The latter certainly does exist in many languages; (8) gives an example from Chukchee (from Spencer 1987).

- (8) *gytli-piri-nin* (cf. *ny-gytle-w* *piri-nin*)
 greedy-take-3sS/3s0 greedy-ADV take-3sS/3s0
 'She greedily took it.'

The crucial difference between adverbs and adjuncts is that adverbs do not form phrases, specifically, they do not take complements or many kinds of modifiers ('She is [greedy for big profits]' vs *'She [greedily for big profits] worked hard'; see Travis in press). Thus, G says nothing about such cases one way or another. Rather, the crucial examples involve those morphemes of time, place, and manner, which are clearly nouns when not incorporated and which can appear with the usual range of nominal specifiers. These do not incorporate. Again, we must narrow in on the natural class about which something worthwhile can be said.³

The claim that locatives do not incorporate also needs clarification, particularly in the light of Mithun's statement to the contrary quoted above. Indeed, there are examples, but to the best of my knowledge they are fairly few and limited. Thus, the sentences in (5) can be described as locative incorporation, but they do not threaten generalization G since the nominal can readily be taken as a direct object of the verb. In fact, essentially all of the examples I have found occur with a small set of motion verbs (like (5)) or posture verbs ('sit', 'stand', etc.). The account offered for (5) can plausibly be extended to all of these cases, since such verbs can take NP objects in other isolating languages (although not necessarily in English). This is quite different from saying that locatives

in general incorporate in any of these languages. One pattern that is never found is examples in which a locative element is required by the verb and appears as a separate word, but its semantic complement is incorporated in the verb.⁴ This is what (7) illustrates. In this case, the idealization is justified by the intuition that what have loosely been called "locatives" in the literature actually must be subdivided into several smaller classes, in ways not yet fully understood.

There is another way locatives incorporate, which confirms generalization G. The full range of locative Ns in the Iroquoian languages do participate in polysynthesis, but by combining with the adpositional element, not the main verb. Examples like (9) from Mohawk form a minimal contrast with the unattested pattern in (7) (from Hewitt 1903):

- (9) *ia'tionte'shennia'te'* *o-hont-ako*
 she-used-whole-strength pre-bush-in
ia-honwa-ia't-onti'
 dir-3FS/3MO-body-threw
 'and with all her might she cast him into the bushes.'

This possibility is expected given that (members of this class of) locatives are the sisters of adpositions and not the sisters of verbs in languages like English. (9) shows again that it is wrong to say that locatives cannot incorporate in the Iroquoian languages; rather we say that the class of elements that a locative can incorporate into is restricted by the structural position it appears in.⁵ Thus, statement G gives significant insight into the basic question of when NI is possible, while a semantically stated alternative is less satisfactory.

3. Extensions of the generalization

The validity of generalization G can be illustrated in more subtle constructions as well, where it interacts with certain distinctions made by current linguistic theory in an interesting way. In addition to deepening the support for G, these constructions show that the term "sister in phrase structure" in the statement of G must refer to relationships at a deep or underlying level of phrase structure, rather than at the surface level. I will consider three additional areas.

3.1 Unaccusative verbs

Intransitive verbs seem to divide into two structural types, a fact that (due largely to Perlmutter) has received much attention in recent years. The distinction is between the unergative verbs, whose subjects are always subjects, and unaccusative verbs, whose surface subjects are actually direct objects at an abstract level of representation. Unergative verbs canonically have agentive or volitional arguments, while unaccusative verbs canonically have patient or theme arguments (although this simple semantic distinction is only a first approximation; see Rosen 1984 for discussion). The difference between the two types of verbs can be seen in many languages, but it is particularly clear in Italian, where, for various reasons, the argument of an unaccusative verb can actually appear in the post-verbal object position on the surface. Thus, contrasts like the following are found (Burzio 1986: 64).

- (10) a. *Interverrà Giovanni a risolvere* UNACCUSATIVE V
will-intervene Giovanni to solve
il problema
the problem
'Giovanni will intervene to solve the problem.'
- b. *??Sperava Giovanni di risolvere il* UNERGATIVE V
hoped Giovanni to solve the
problema
problem
'Giovanni hoped to solve the problem.' (OK is *Giovanni sperava di ...*)

Now, generalization G leads us to expect that these two verb classes could work differently in Noun Incorporating languages. (11) shows the structures of sentences with the two types of verbs.

- (11) a. UNACCUSATIVE VERBS UNERGATIVE VERBS
-
- ```

graph TD
 S1[S] --- NP1[NP]
 S1 --- AUX1[AUX]
 S1 --- VP1[VP]
 NP1 --- c[c]
 VP1 --- V1[V]
 V1 --- spill[spill]
 VP1 --- NP2[NP]
 NP2 --- beans[beans]

 S2[S] --- NP3[NP]
 S2 --- AUX2[AUX]
 S2 --- VP2[VP]
 NP3 --- louse[louse]
 VP2 --- V2[V]
 V2 --- crawl[crawl]

```

Since the unaccusative verbs have an argument as a sister of the verb at an abstract level of structure, this argument could potentially undergo Noun Incorporation. This would never be possible with the argument of an unergative verb, however, since this argument is never a sister of the verb. This correctly describes the situation found in the Iroquoian languages; Woodbury (1975) gives the following from Onondaga.

- (12) a. *Ka-hsaheʔt-ahi-hw-i*  
they-bean-spill-CAUS-ASP  
'The beans spilled.'
- b. *\*H-ate-tsiʔkti-ʔse:-ʔ*  
it-REFL-louse-draf-ASP  
'The louse crawled.'
- OK: *H-ate-ʔse:-ʔ neʔ otsiʔktaʔ*  
it-REFL-drag-ASP DET louse  
'The louse crawled.' (no NI)

The same kind of contrast is observable in Southern Tiwa (Allen, Gardiner, and Frantz 1984). Indeed, this gives us an account of what Mithun claims (1984: 875) to be the most common type of NI language; she writes: "If a language incorporates only two types of arguments, they will be *patients* of transitive and *intransitive* V's [emphasis added – MCB]." Again, there is striking parallelism between the Ns that can incorporate in polysynthetic languages and the Ns which are structural objects in English and Italian.

### 3.2 Instruments and benefactives

Continuing the exploration of different constructions, let us investigate briefly the topic of oblique nominals. Mithun states that instrumental Ns may be incorporated in some languages. This occurs in, for example, Nahuatl (from Merlan 1976).

- (13) *Yaʔ ki-kočillo-teteʔki panci*  
he he/it-knife-cut bread  
'He cut the bread with a knife.'

Certain other oblique semantic roles are conspicuously absent from the list of possible NIs, however. The most obvious are the benefactive and dative roles, which are not incorporated in any polysynthetic language.<sup>6</sup>

- (14) a. \**Wa<sup>2</sup>-khe-yat-wir-ahnin<sup>2</sup>-0* (Tuscarora; Williams 1976)  
 PST-1sS/30-REFL-child-sell-ASP-APPL  
 'I sold him to the children.' (OK as 'I sold the children to him.')
- b. \**Ta-hliawra-( 'u'u)-wia-ban* (Southern Tiwa, Allen, I:her/her-woman-baby-give-PAST Gardiner, and Frantz 1984)  
 'I gave the woman it (/the baby).'  
 (OK is *Ta- 'u'u-wia-ban hliawra-de*  
 I:her-baby-give-PAST woman-SUF, without NI of the goal)

Superficially, these facts seem problematic for our generalization, since both instruments and benefactives appear as the objects of prepositions in English (e. g., *Picasso made this figurine for Mary with a penknife.*). If anything, the benefactive looks more like the object of the verb than the instrument does, given the existence of dative shift (e. g., 'Picasso made Mary a figurine with this penknife'). Thus, (13) and (14) are the opposite of what we might expect.

Appearances can be deceiving, however. The discussion of intransitive verbs makes it clear that NI can be sensitive to underlying structural relationships, rather than just to superficial ones. Now, it is standard in Government and Binding theory to assume that some prepositions do not exist in underlying structure, but are inserted later. An example is the semantically empty *of* in (15 b) (cf. (15 a)).

- (15) a. *I watched the Romans destroy Carthage*  
 b. *I witnessed the Romans' destruction of Carthage*

Other prepositions might exist in underlying structure but be deleted on the surface. This occurs in the traditional transformational account of dative shift, in which sentences like (16 b) are derived from sources like (16 a).

- (16) a. *Picasso made a figurine for Mary*  
 b. *Picasso made Mary a figurine*

In fact, Baker (1988 b) argues (largely from applicative constructions in Chichewa) that benefactives are always sisters of adpositional elements underlyingly, while instrumentals are sisters of the verb. English com-

pounds give some evidence to this effect; note that instruments but not benefactives can productively appear in deverbal compounds.

- (17) a. *hand-made scarves, fork-split English muffins, laser-cut diamonds*  
 b. \**child-made scarves* (i. e., made for children), \**cat-prepared meat* (prepared for cats), \**church-given contributions*

This is easily explained if one assumes that prepositional phrases cannot appear in a compound word in English. This restriction makes it impossible to have a benefactive in a compound, since benefactives must be full PPs underlyingly. True, in full sentences the *for* may disappear as a result of the process that effects dative shift, but this cannot rescue the compound. Instruments, on the other hand, are possible in compounds because the P is not semantically necessary and need not appear underlyingly. *With* does normally appear with instrumental phrases in full sentences in English as a result of a preposition insertion rule of some kind,<sup>7</sup> but this need not apply in the compound.

Now, if these same underlying relationships are present in the polysynthetic languages, then it is not surprising that instrumentals can incorporate in some of them, but benefactives cannot. This is exactly what generalization G predicts, since instruments are structurally like objects at the relevant level, while benefactives are structurally parallel to the locatives discussed in the preceding section (see (14 b') below). Again, parallels between English and polysynthetic languages appear.

### 3.3 Causative verbs

Finally consider the following causative sentence in Southern Tiwa (Allen – Gardiner – Frantz 1984).

- (18) *Ti-seuan-p'akhu-kumia-'am-ban wisi*  
 I/he-man-bread-sell-CAUS-PAST two  
*te-khaba-<sup>2</sup>i*  
 I/them-bake-SUBORD  
 'I made the man sell the two breads that I baked.'

Apart from the causative morphology, (18) looks, on the surface, exactly like the impossible structure with a goal NP given in (14 b). Each has