Advances in Functional Linguistics
Studies in Functional and Structural Linguistics (SFSL)

Taking the broadest and most general definitions of the terms functional and structural, this series aims to present linguistic and interdisciplinary research that relates language structure — at any level of analysis from phonology to discourse — to broader functional considerations, whether cognitive, communicative, pragmatic or sociocultural. Preference will be given to studies that focus on data from actual discourse, whether speech, writing or other nonvocal medium.

The series was formerly known as Linguistic & Literary Studies in Eastern Europe (LLSEE).

Founding Editor
John Odmark

Honorary Editors
Eva Hajičová
Charles University

Petr Sgall
Charles University

General Editors
Yishai Tobin
Ben-Gurion University of the Negev

Ellen Contini-Morava
University of Virginia

Editorial Board
Alexandra Y. Aikhenvald
La Trobe University

Jim Miller
University of Auckland

Joan Bybee
University of New Mexico

Marianne Mithun
University of California, at Santa Barbara

Nicholas Evans
University of Melbourne

Lawrence J. Raphael
Emeritus CUNY

Víctor A. Friedman
University of Chicago

Olga Mišeska Tomić
Leiden University

Anatoly Liberman
University of Minnesota

Olga T. Yokoyama
UCLA

James A. Matisoff
University of California, Berkeley

Volume 57
Advances in Functional Linguistics: Columbia School beyond its origins
Edited by Joseph Davis, Radmila J. Gorup and Nancy Stern

Advances in
Functional Linguistics
Columbia School beyond its origins

Edited by
Joseph Davis
The City College of New York
Radmila J. Gorup
Columbia University
Nancy Stern
The City College of New York

John Benjamins Publishing Company
Amsterdam/Philadelphia

Library of Congress Cataloging-in-Publication Data


p. cm. (Studies in Functional and Structural Linguistics, ISSN 0165–7712 ; v. 57)
Includes bibliographical references and indexes.
I. Functionalism (Linguistics)

P147.A384 2006
410.1/8–dc22 2006043042
isbn 90 272 1566 9 (Hb; alk. paper)

© 2006 – 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 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA
Table of contents

List of contributors ix

INTRODUCTION
Consistency and change in Columbia School linguistics
Joseph Davix 1

Linguistic theory

CHAPTER 1
Columbia School and Saussure’s langue
Wallis Reid 17

CHAPTER 2
Diver’s Theory
Alan Huffman 41

Phonology

CHAPTER 3
Phonology as human behavior:
Inflectional systems in English
Yishai Tobin 63

CHAPTER 4
Phonological processes of Japanese based on the theory
of phonology as human behavior
Yishai Tobin and Haruko Miyakoda 87

CHAPTER 5
Phonology as human behavior:
A combinatory phonology of Byelorussian
Igor Dreer 107
CHAPTER 6
Phonology as human behavior:
The case of Peninsular Spanish
Adriaan Dekker and Bob de Jonge

CHAPTER 7
Functional motivations for the sound patterns of
English non-lexical interjections
Gina Joue and Nikolina Collier

CHAPTER 8
Phonology without the phoneme
Joseph Davis

Grammar and lexicon

CHAPTER 9
Tell me about yourself:
A unified account of English -self pronouns
Nancy Stern

CHAPTER 10
Se without deaxis
Radmila J. Gorup

CHAPTER 11
The difference between zero and nothing:
Swahili noun class prefixes 5 and 9/10
Ellen Contini-Morava

CHAPTER 12
A semantic analysis of the Swahili suffix li
Robert A. Leonard and Wendy Saliba

CHAPTER 13
The structure of the Japanese inferential system:
A functional analysis of daroo, rashii, soo-da, and yooda
Hidemi Sugi Riggs
CHAPTER 14
Structuring cues of conjunctive *yet, but, and still*
A monosemic approach

*Charlene Crupi*

263

**Beyond language**

CHAPTER 15
The case for articulatory gestures – not sounds – as the physical
embodiment of speech signs

*Thomas Eccardt*

283

CHAPTER 16
Meaning in nonlinguistic systems:
Observations, remarks, and hypotheses on food, architecture,
and honor in Kenya

*Robert A. Leonard*

309

Index of names

335

Subject index

339
List of Contributors

Nikolinka Collier
Department of Computer Science
University College Dublin
Dublin 4
Ireland
niki.collier@gmail.com

Ellen Contini-Morava
P.O.- Box 4120
Department of Anthropology
University of Virginia
Charlottesville, VA  22906 - 4120
USA
elc9j@virginia.edu

Charlene Crupi
155 West Lake Road
Lake Paupac Club
Greentown, PA  18426
USA
ccrupi@rci.rutgers.edu

Joseph Davis
The City College of New York
School of Education, NA 6/207
New York, NY  10031
USA
davis@ccny.cuny.edu

Igor Dreer
Department of Foreign Literatures
and Linguistics
Ben-Gurion University of the Negev
P.O. Box 653
84105 Be’er Sheva
Israel
dreer@bgu.ac.il

Thomas Eccardt
455 East 14th St. #3-C
New York, NY  10009
USA
teccardt@hotmail.com

Radmila J. Gorup
Columbia University
Department of Slavic Languages
and Literatures
Hamilton Hall
New York, NY  10027
USA
gjg26@columbia.edu

Alan Huffman
Program in Linguistics
The Graduate Center
City University of New York
365 Fifth Avenue
New York, NY  10016
USA
jh620@columbia.edu

Adriaan Dekker
Opaalweg 117
3523 RM Utrecht
Netherlands
adriaan.dekker@tiscali.nl
Introduction

Consistency and change in Columbia School linguistics

Joseph Davis
The City College of New York

1. This volume and its predecessors

This is the fourth in a series of volumes that has emerged from a series of linguistics conferences organized by the Columbia School. This volume thus adds to an existing body of research, but it also develops the series in a way that the others have not. This is the first volume of the series – indeed, to our knowledge, the first anthology in existence – to focus exclusively on Columbia School work. This is not because papers from other traditions were rejected, nor because the present editors doubt the merit of dialogue with other schools. Rather, what happened is that a freestanding Columbia School volume materialized out of the seventh (2002) conference, and the editors recognized that as fortuitous. It afforded us an opportunity to make a statement about Columbia School’s maturity even while shouldering the more important responsibility of contributing to the School’s growth and change and, thereby, to the larger field’s evolving understanding of human language. Being thus theoretically focused, the volume advances Columbia School (CS) work by re-examining its most fundamental issues, contributing new analyses, and testing the applicability of lessons learned on new frontiers. We believe that this focus is appropriate in light of the increased attention that CS has received in the years since the conferences began, most noticeably in such full-length works as Reid (1991), Huffman (1997), Tobin (1997), and Contini-Morava and Tobin (2000).

The editors of the three conference-related volumes that precede this one (Contini-Morava and Goldberg 1995; Reid, Otheguy, and Stern 2002; Contini-Morava, Kirsner, and Rodríguez-Bachiller 2004) all made concerted and highly successful efforts to include papers from other functionalist perspectives and to explore both differences and commonalities between those perspectives and CS itself. In hindsight, those volumes can now be seen to have led quite naturally to the narrower focus of the present volume. In reviewing those documents today, one can scarcely
miss the growing confidence of an intellectual movement first reaching out to peers who might be amenable to dialogue on the basis of some shared outsiders’ interest, then baldly stating what exactly makes us different from the rest, and finally engaging in full-fledged debate about what separates and unites us.

As one measure of that change, consider the decreasing attention given to generative grammar. In her introduction to the first volume, Ellen Contini-Morava (1995) dedicates the first thirteen of twenty-eight pages to differentiating the functionalist perspectives in that volume from generative grammar and syntactic theory. In his introduction to the second volume, Wallis Reid (2002) makes just three brief mentions of generative grammar, calling its treatment of semantics a “guarded entry” into the realm of meaning (p. ix), citing the secondary status it accords to actual language use (p. xiii), and admitting that “a theory of linguistic behavior [such as CS] is unfashionable in the Chomskyan era” (p. xx). Robert S. Kirsner’s (2004) introduction to the third volume, our immediate predecessor, nearly ignores generative grammar altogether, maintaining that a comparison between it and CS “might be of little use, because of the tremendous differences in the two schools’ orienting premises” (p. 2).

Anyone interested in the big questions in linguistics could scarcely do better than to read those three introductions. It will be useful to review them here, both to show those familiar with those volumes how this volume fits into the developing series, and to help to orient newcomers to the overall theoretical perspective that will be found in the pages to follow. Of course, no summary such as this can do justice to the actual documents themselves. Readers seeking an excellent article-length introduction to CS are referred as well to Huffman (2001).

1.1 Meaning as explanation: Advances in linguistic sign theory (1995)

Contini-Morava (1995), introducing the first volume of the series, identifies “two language-external, controlling principles” which unite the various “sign based theories” represented in that volume: Columbia School, Guillaumean, Jakobsonian, Cognitive Grammar, and others. These controlling principles are communication, embodied in the linguistic sign, and human psychology (p. 2).

Contini-Morava contrasts these sign-based theories with generative grammar in terms of definition of data and conception of meaning. Whereas generative grammar takes as data “native speakers’ judgments about sentence structure and well-formedness,” sign-based theories take the distribution of linguistic forms as the phenomenon to be explained. And while generative grammar grapples with “the ‘creativity’ of language, defined as a speaker’s ability to produce and understand an infinite number of formally distinct sentences,” sign-based theories explore how we “account for the fact that human beings are able to produce and
infer an infinite number of novel messages from a finite number of signs” (p. 3). As for their treatment of semantics, “all formal theories subscribe to some version of compositionality,” in which “each sentence has a determinate number of meanings,” “calculated on the basis of the meanings of the lexical items in the sentence together with a set of semantic rules that specify how to interpret the syntactic configurations supplied by the syntactic rules” (p. 5). By contrast, CS sees the relation between signaled linguistic meaning and inferred communicative message as one of “contribution” rather than composition, in that “the meanings of individual linguistic signs are not always readily identifiable as fractional components of a notional whole” but may instead merely provide hints for interpreting an utterance in context (p. 6). Contini-Morava then summarizes the ways that sign-based theories handle phenomena regarded elsewhere as syntactic (e.g., word order, agreement, and government).

According to Contini-Morava, sign-based theories, including CS, accept Saussure’s langue, with its signs defined by oppositions of value, as distinct from parole, or individual acts of speech. (On this point, debate continues; see below.) CS diverges from Saussure in including substance as well as value in its meanings, in distinguishing grammar and lexicon, and in invoking the Human Factor, the fact that language-users “are human, hence have particular abilities, biases, and limitations that affect the nature and patterning of linguistic forms” (p. 17).

CS differs from other functionalist theories in excluding parts of speech as linguistic categories, and in going beyond the sentence for analytical validation, thus creating the need for quantitative validation across large stretches of discourse. CS also differs from other functionalists in the particulars of how it represents grammatical meaning. CS posits semantic substances which are exhaustively divided into a finite number of grammatical meanings, each with a signal that is typically “satellite” to a lexical item (For example, Sanskrit has three signaled Number meanings – one, two, more – that give information about the referent of the lexical item to which their signals are attached). CS invokes both oppositions of exclusion (English, Time past and non-past) and of inclusion (Homeric Greek, Number two and more than one), and allows for the “interlock” of two or more grammatical systems simultaneously using the same morphology (e.g., Discourse Referent, Number, and Time in English verbal -s). CS is avowedly not universalist in its aims, limiting its claims to just those languages investigated and neither assuming nor seeking wider generalities (though not ignoring cross-linguistic parallels when they do emerge from analysis). CS does not assume a theory of markedness in semantic structure. Finally, up to the time of Contini-Morava (1995), CS linguists regularly invoked the communicative strategy as “a routinized exploitation of a given meaning, so that it is regularly used to suggest/infer a particular type of message” (p. 19, but see below).
1.1.1 Connections to the present volume

Although at least in some sense – perhaps, one might argue, for all practical purposes – CS is indisputably “sign based,” as Contini-Morava calls it, the exact aptness of that characterization remains the subject of debate. Beyond the perhaps niggling question of precisely which works, published and unpublished, authentically represent CS, the matter has unavoidably to do with Saussure (1916 [1972]). It has to do with Saussure’s conception of the signe, a conception known to us only through the posthumous *Cours de Linguistique Générale*. And it has by extension to do with Saussure’s conception of langue, an equally problematic matter. Diver (1974) unequivocally places himself in a line descending from Saussure, and in that document he explicitly uses the terms signe and langue. At least at first glance, he appears to be aligning himself with Saussure. By contrast, Diver (1995: 87) disparages the “philosophical implications” of Saussure’s use of *substance* and *valeur*, studiously avoids using the word *sign*, and utterly refutes the idealization of a language system shared by all members of a speech community. Davis (2004a) analyzes these two documents by Diver and downplays Saussure’s influence on at least the later Diver. In the present volume, Wallis Reid responds, attempting to redress the balance, as it were, and give Saussure his due. It is probably the case that Davis and Reid agree on the facts – that Diver does not share exactly Saussure’s conception of signe and langue – and disagree merely on the relative question of whether Diver’s debt to Saussure is strong enough to merit characterizing Diver’s linguistics as being particularly “Saussurean” (cf. Kirsner 1993; Otheguy 2002: 374).

Alan Huffman, in the present volume, has a different take from Contini-Morava’s on what distinguishes CS from other theories. His paper is a critical reading of the final statement by Diver (1995), a paper titled simply “Theory.” Huffman argues that the essential idea of that landmark paper has not been fully appreciated even within CS. Diver’s main point, says Huffman, was that linguistic theory must be far more aposterioristic and minimalist than has ever been the case. That position is consonant with Diver’s longstanding criticism of linguistics for its continuing reliance on inherited categories that pretend to solve large problems even before any small problems have achieved analytical success. Huffman maintains that this austere position is what differentiates CS not only from traditional and generative grammar but even from other sign-based and functionalist schools. Finally, Huffman examines the relationship between Diver’s theory and philosophy of science.

1.2 Signal, meaning, and message: Perspectives on sign-based linguistics (2002)

In his introduction to the second volume, Reid (2002) assumes familiarity with Contini-Morava (1995) and brings to the fore certain points raised by chapters in his vol-
Introduction: Consistency and change in Columbia School linguistics

Reid provides an excellent and succinct statement of what he views as the “true import” of the phrase “meaning as explanation” in the title of the first volume:

The place of meaning in linguistic theory has been one of the most contentious in the field. Its treatment has ranged from indefinite postponement by Bloomfield, guarded entry by Chomsky, to the central place it occupies in cognitive grammar. On one point, however, there has been general agreement: a major goal of any linguistic theory is to offer an explicit account of meaning in language. Columbia School turns the tables here and makes linguistic meaning the *explanans* in linguistic theory rather than the *explanandum*. Systemic meaning functions to explain the distribution of linguistic forms rather than being itself the object of explanation. (Reid 2002: ix)

By narrowing the definition of *meaning* to just that which is systematized in language, and thereby rejecting the goal of explaining meaning in the larger sense, CS is able to define a different problem to be solved, a radically pre-theoretical problem having to do ultimately with a physical phenomenon in the world; in the words of Diver (1995: 49): “discovering the motivation for the particular form taken by sound waves produced by the human vocal apparatus.” Reid’s wording (p. ix), “to explain the distribution of linguistic forms,” amounts to essentially the same thing but at a higher level of abstraction and a level closer to the one where CS linguists usually work.

The implications of such an “epistemologically austere position” are far-reaching (p. xiii). Having a new problem means that the old solutions may or may not help you. CS has effectively adopted Saussure’s “radical anti-nomenclaturism,” his dictum that language is a principle of classification unto itself. “The categories of linguistic meaning and structure are both language internal and language particular; they do not come from outside” (p. xiii; cf. Otheguy 2002).

Clearly, if CS linguists cannot rely on the tradition for guidance, then the burden on us to validate the categories we come up with is vastly increased. “When the distribution of linguistic forms in authentic discourse becomes the explicit object of explanation, the procedures for data collection and analysis assume a heightened importance” (p. xv). That is why we bolster our hypotheses with predictions about the distribution of forms across large stretches of text and then test those predictions quantitatively.

Given that CS has so thoroughly rejected mainstream constructs, what can it hope to contribute to the “intellectual marketplace” (Kirsner 2002: 341 *et passim*)? Reid concurs with Kirsner:

The answer . . . lies in [Columbia School’s] therapeutic function and in its exquisite empirical methodology. By proposing analytical problems and solutions which could simply not be conceived in other frameworks, it serves as a corrective force
Advances in Functional Linguistics

in the field as a whole. Columbia School’s skepticism towards all received categories of analysis forces scholars who take such categories for granted to reconsider, however unwillingly, exactly what they are doing. Finally, its concern with both qualitative and quantitative validation liberates the linguist from the conceptual and empirical shackles of the sentence-based perspective still held by most other schools of linguistics. (Reid 2002: xx-xxi)

In some ways, Reid, Otheguy, and Stern (2002) is the most thoroughly Columbia School of the series to date. Its title, Signal, Meaning, and Message, is explicitly Columbia School. Its introduction is focused on CS. And it contains, by our reckoning, just three papers (out of thirteen) from outside the CS tradition. Reid does, however, reach out to Cognitive Grammar, pointing out that three CS papers in the volume employ cognitive rationales as at least part of their account for the distribution of linguistic forms. That is, “these two modes of explanation” – the cognitive and the communicative, to anticipate the title of the third volume – “need not be mutually exclusive” (p. xi).

1.2.1 Connections to the present volume
Ellen Contini-Morava’s contribution to the present volume serves as a model for how an analyst armed with such “skepticism towards all received categories,” can break free of received paradigms. She wrestles with the always challenging problem of zero (Ø) as a potential linguistic signal, all the more challenging outside the confines of a closed system such as the familiar European case paradigm. Three Swahili noun classes have prefixes (also signaling Number) that are traditionally analyzed as involving Ø as an allomorph. Contini-Morava argues, however, that only one of those classes actually has a Ø morpheme, the others instead having no prefix at all. The argument hinges on the lack of distinct plural forms for the latter classes and on morphophonemic analysis of nouns of those classes.

1.3 Cognitive and communicative approaches to linguistic analysis (2004)

The third volume in the series is to some extent a concerted dialogue between Cognitive Grammar (CG) and Columbia School, a dialogue which Robert S. Kirsner has pursued for some time (1989, 1993, 2002) and which Contini-Morava (1995) and Reid (2002) took up as well.

Kirsner’s theme, in his introduction (2004) to the third volume, is the debate between CG and CS as to where to draw the line marking the limit of the linguist’s responsibility. As we saw in Reid (2002), CS takes the “austere” (p. xiii) or modest (p. xx) or even “reductionist” (p. xix) position that it is accountable only for the distribution of linguistic forms in various languages. CS does not undertake to dis-
cover linguistic universals. Nor does CS take responsibility for explaining meaning or communication in a larger sense. (It is worth pointing out that Kirsner ignores here Diver’s [1995] still more extreme claim of responsibility only for successful analyses accomplished one by one, or, in other words, for beginning with “zero responsibility.”)

Kirsner is justifiably challenged and intrigued by Ronald Langacker’s (2004) thoughtful critique of CS in the lead chapter of the volume. Langacker accepts the sparse CS meaning as a legitimate linguistic unit, but he pushes CS to accept responsibility as well for facts of usage which show evidence of “psychological entrenchment and conventionalization within a speech community.” CS had in fact operated for years with a construct intended to do just that, yet without ever fully according it membership in the linguistic system. This construct was the communicative strategy, whose development in CS thought Reid (1995) so carefully traced, and which Reid (2004) continues to rely upon, yet without, as Kirsner notes (p. 6), making strategy part of the “communicative instrument” itself. Like Kirsner (1989) before him, Langacker wants CS to pay more attention to strategies, to take responsibility for them, to consider them “part of a language” (p. 49). He wants us to do more than show “how the abstract meaning posited can be reconciled with each attested use via some interpretive process” by also showing “how the specific, actual range of uses follows inexorably from the abstract characterization, so that this alone would exhaust the information a speaker needs in order to use the element correctly (i.e. in accordance with the conventional expectations of the speech community)” (p. 46). Within the very same volume, meanwhile, Davis (2004b) levels a CS-internal attack against strategy even as an analytical tool kept outside the linguistic system proper. Far from wanting to exalt its status, Davis argues for abandoning it altogether.

Given this dispute, Kirsner’s concern is that:

by jettisoning the concept of strategy, Davis may be further restricting the scope of phenomena for which a Columbia School linguistic analysis would claim responsibility, thereby making the Columbia School approach more vulnerable to Langacker’s criticism that it attempts to explain too little about Language to be interesting. (Kirsner 2004: 9)

Kirsner muses (p. 6), “How immediately ‘useful’ should a linguistic theory be and in what ways?” Or, in Langacker’s words (quoted in Kirsner 2004: 3–4), “The more narrowly linguists define their own work, the less they will contribute to the global enterprise of figuring out how language works.” It is probably proper, and unavoidable, then, that Kirsner, having so dutifully contributed to and facilitated the debate, should conclude thus: “At this point it is up to the reader to judge how pro-
Advances in Functional Linguistics

ductive these two approaches to linguistics – the cognitive and the communicative – ultimately are” (p. 7).

1.3.1 Connections to the present volume
Having myself stoked the flames in the debate over strategies, and having now the vehicle of this introduction at my disposal, I might offer a suggestion tending toward conciliation. It is true that Diver (1995) takes an extremely austere position with regard to the linguist’s area of responsibility, and that others in CS (mainly Diver’s students) have been scarcely more moderate. Yet one cannot but suspect that the real reason CS linguists preoccupy ourselves with the apparently tedious problem of where certain forms show up in texts, is that we too are ultimately interested in “the global enterprise of figuring out how language works.” We go about it the way we do, setting such modest goals, because we enjoy analytical success, and Diver convinced us that, at this point in the development of the field, we can hope to achieve analytical success only by tackling very small problems. In this way, we can indeed contribute to the global enterprise, albeit only incrementally for the present. Maybe laying a stone now and then is as constructive as imagining the finished cathedral. At least, in Kirsner’s words, CS is not “doomed to remain eternally programmatic, making perhaps attractive majestic and ambitious claims about the mind, but never actually delivering on its promises” (p. 7).

It should not be overlooked that Langacker (2004: 49) speaks of a “gradation”: “A linguistic structure is reasonably considered part of a ‘language’ to the extent that speakers learn it and can use it as a prepackaged unit (an established cognitive routine), and can count on other speakers knowing it as well” (emphasis added, JD). Langacker would replace the meaning-message distinction with a meaning-message continuum. Factor in Langacker’s recognition that much of the argument hinges on “the terminological issue of how we choose to use terms such as ‘language’” (p. 47), and one may begin to suspect that, deep down, the main thing we disagree on is how we prefer to spend our time: accounting for the distribution of forms or accounting too for what people choose to talk about.

Radmila Gorup, in the present volume, furnishes the first published CS analysis that by design does not appeal to the concept of strategy. The distributional facts of Serbo-Croatian se resemble the uses that led earlier CS analysts to posit a system of Deixis, or Attention, to account for referent-finding strategies (low Attention to find an obvious referent, a reflexive, or to suggest that finding the referent isn’t important, the impersonal). Gorup dispenses with a system of Deixis, showing that the distribution of se in discourse can be accounted for purely on the basis of the sparse meaning central Focus on a participant (a meaning shared with the nominative case). Se’s absence of other meanings such as Number, Gender, and Person, accounts for its distributional differences vis-à-vis other pronominal clitics.
Another criticism that Langacker has of CS is our reluctance, or perhaps even refusal, to entertain the possibility of the grammatical construction, or a sign made up of signs (Otheguy 1995). It is true that we have strenuously resisted positing constructions where others have been happy to see them, but that is because we have found that, by breaking supposed constructions up into their constituent parts, we get a closer fit with the facts (Huffman 2002).

Nancy Stern, in the present volume, offers an analysis of the English -self pronouns that simply avoids the question of whether each of them is to be analyzed as one signal or two (or more). Few will dispute that her semantic hypothesis is insightful, accounting for all uses of the forms, including reflexive, emphatic, logophoric, and other so-called exceptional uses. Stern relies neither on the traditional category of reflexive (a rejection going back to García 1975) nor on the earlier CS substance of Deixis or Attention. While a typical CS treatment might have required a semantic analysis based more closely on a strict morphemic cutting (i.e., himself = him + self), her analysis shows that something interesting and useful can be learned through a CS approach even if one is unconcerned with the surgical avoidance of construction.

2. Consistency in Columbia School linguistics

Amidst all the inter- and intramural dialogue, fruitful as it is, what might go unacknowledged is the large degree of consistency that has characterized CS theory and practice over the years. Now consistency may or may not be a good thing; one can be consistently wrong either through obstinacy or through inertia. But as we present this fourth anthology of CS work, aiming to take “Columbia School beyond its origins,” it behooves us to assess what has remained constant before we assert that something constitutes change.

The analyses presented in this volume will look much like CS (or form-content) analyses of the 1970s. And that is because the program really has remained quite consistent in both theory and practice. The seminal CS work, Diver’s (1969) analysis of the Homeric Greek verb, already:
- took observable sound as the phenomenon to be explained;
- rejected syntax;
- adopted communication, with the meaningful signal, as an orienting principle;
- divided the overall task into two parts, phonology (distribution within the linguistic signal), and grammar and lexicon (distribution of the linguistic signal), and further distinguished between the latter two on the basis of systematicness of organization;
- distinguished between signaled meaning and inferred message;
– posited a particular (not universal) grammatical system of a semantic substance exhaustively divided into oppositional meanings, each with its signal;
– recognized the interlock of grammatical systems;
– employed authentic discourse as the source of data;
– correlated the appearance of signals with other features of the text that are independent of the hypothesis being tested, thus employing quantitative validation; and
– used an understanding of the overall message in a passage to validate the contribution of a hypothesized meaning to that message.

The main thing absent from that first paper is explicit mention of the Human Factor, which may, however, be implicit, and which certainly is strongly implicit in Diver (1974) and becomes fully explicit in Diver (1975). The only idea that has definitively been discarded from Diver (1969) is the curious position taken then that messages are somehow observable. One major innovation since 1969 was, as we have seen, the communicative strategy, but that construct has never quite made its way fully into the theoretical apparatus. A more theoretically minor, but nevertheless highly visible, innovation that has since come and gone is the use of the chi-square test of significance on quantitative data (Davis 2002b).

In the present volume, a representative of a CS-type grammatical analysis is Hidemi Sugi Riggs’s treatment of the Japanese auxiliaries commonly translated as ‘probably’, ‘seem’, ‘I have a feeling that’, etc. These have traditionally been identified as epistemic markers or evidentials. After showing the flaws in those analyses, Riggs presents a system of oppositions involving two semantic parameters, Inference RELEVANT/NOT RELEVANT and Focus (emphasis) HIGH/LOW, to account for the observed distribution in authentic discourse.

Consistency is likewise to be found in CS phonology: The papers in this volume will bear a striking resemblance to the seminal CS phonological work, Diver (1979), which already:
– took as its task to account for the distribution of articulations within the lexicon;
– adopted communication, phonetics, and the Human Factor as orientations;
– employed quantitative methods; and
– took the position that a purely descriptive linguistics, without explanation, is not possible.

Adriaan Dekker and Bob de Jonge, in the present volume, continue along the vein established by Diver (1979) in examining what appears to be an exception to the typical distribution of phonological units in lexicon. In Peninsular Spanish, initial \( k \) is observed more frequently than would be expected on the basis of general
phenomena that have earlier been shown to influence distribution (number of articulators required, plosiveness, adroitness of articulator, visibility). The authors suggest that particularities of the Spanish phonological system may account for the aberration.

Thomas Eccardt’s contribution to the present volume, a paper on semiotic theory, maintains that Diver was among the few who have treated articulatory gesture as the substance of their phonological units. He takes issue, however, with Diver’s insistence that sound is the only true observable in language.

Readers will find, we believe, that the papers in the present volume remain, by and large, consistent with all the longstanding characteristics of CS work. Whether such consistency is good or bad for linguistics will have to wait for posterity to judge: whether it is due to obstinacy or inertia, or whether instead it suggests that CS, through all these years, has been close enough to the right path to give analysts enough of a sense of success that they have been willing to persevere on what they see as a promising program of research in spite of its being a perhaps narrow path with an invisible endpoint.

3. Change in Columbia School linguistics

To say that CS has been remarkably consistent over the years is not to say that it has been static. New analyses have continued to come out, and theoretical issues have been explored. In the process, the infamously modest aims of CS have grown modestly less modest. Several papers in the present volume illustrate this willingness – or dare we say eagerness – to confront the possibility of change.

Phonology in this volume gets liberated from the standard dictionary and from the confines of the monosyllabic stem. Yishai Tobin builds on his record of advancing CS phonology, or Phonology as Human Behavior (PHB), extending it into the synchronic and diachronic analysis of English inflection. What Tobin calls the fundamental axiom of PHB is that language represents a struggle between the needs of maximum communication and minimal effort. He presents evidence that, synchronically, inflectional morphology, being both functional and frequent, tends to be composed of phonological units that require little effort on the part of the speaker. He shows that, diachronically, this tendency increases over time.

Tobin collaborates with Haruko Miyakoda in a paper that examines the phonological processes (substitution, epenthesis, deletion) observed in Japanese speech errors and loan words from the point of view of CS phonology. They present evidence that the various communicative forces found within different word positions have an influence on how and where the phonological processes occur.
Gina Joue and Nikolinka Collier examine two speech corpora for a relation between the sound patterns of nonlexical interjections (e.g., \textit{m-hm}) and their functional realization in discourse. The degree of what they call the phonological markedness of an interjection correlates with its use in dynamic vs. static interaction. Multisyllabic interjections correlate with dynamic interaction, as opposed to monosyllabics with static interaction. The analysis provides support for the view that the function of interjections depends on a combination of their position in conversational turns, their denotation, and context.

In my own paper, I argue that the classical structural phoneme never properly earned its place into CS phonology and indeed is not required by theoretical principles. The phoneme was indeed a theoretical construct for Diver, but Diver glossed over well-known obstacles in phonemic analysis, obstacles that preclude the degree of analytical success he typically demanded. Diver evidently believed that the phoneme was required by CS’s communicative orientation, but I show that it is not. Linguistic signals or lexemes can be segmented into CS phonological units without insisting that these equate to the distributionally defined phoneme.

It has been noted that CS classically posits semantic substances exhaustively divided into relational meanings and has not typically made use of the notion of markedness.\footnote{Robert Leonard and Wendy Saliba, however, propose that the Swahili suffix -\textit{li} has a meaning which is neither absolute (e.g., patient) nor relational to other signals in a closed system (cf., e.g., Huffman 1997). Rather, -\textit{li} is auxiliary to such a system (cf. Goldberg 1995), yet having a sometimes dramatic impact on the interpretation of the opposition between the relative meanings of that system. -\textit{li} functions as a signal that a non-high controller of an event (e.g., direct object) should be interpreted as having more control than would be the case if -\textit{li} were absent.}

One can wonder why CS has historically shied away from delving into lexicon. Perhaps it is because of the supposedly more complex, less structured network of oppositions lurking there, compared with the relatively one-dimensional gradients found in a typical grammatical system. This would be consistent with Diver’s (1995) advice to solve simple problems first. Though there have been interesting forays, e.g., de Jonge (1993) and Tobin (1995), for the most part, lexicon remains for us a vast unexplored terrain. Charlene Crupi takes us there in her analysis of English \textit{yet} as it semantically opposes \textit{but} and \textit{still}. Yet signals significant contrast: that two chunks of information surrounding \textit{yet} differ in ways that are deemed thematically important. Unlike \textit{yet}, \textit{but} is thematically asymmetrical, and \textit{still} is resumptive. Crupi’s line of attack on \textit{yet} is to oppose it to just those uses of other lexical items, \textit{but} and \textit{still}, which effectively commute with \textit{yet}, leaving aside the more irrelevant (for her purposes) uses of those words. Like the paper by Stern, this one makes extensive use of electronic corpora, also atypical in CS work to date.
Robert Leonard’s paper on “Meaning in Nonlinguistic Systems: Observations, remarks, and hypotheses on food, architecture, and honor in Kenya” concerns nonlinguistic systematic meaning. We offer the paper as a first test of the idea (cf. Huffman, this volume) that Diver’s linguistic theory has a larger epistemological relevance. Leonard offers an anthropological analysis of behavior involving the use of space and food. In this paper, as in CS grammatical analysis, observations of physical phenomena lead to hypotheses which are then tested against further observations. Here as in CS traditionally, the hypotheses are relational and exhaustively divide up a scale. And here as there, the connection between observation and significance is at least semi-arbitrary. We trust this excursion “beyond language”—tentative as it may be—will prove to be provocative.

None of the papers in this volume, of course, can be thought of as the final word on its topic. Some are more tentative, some more grounded in precedent. Many of the ideas presented here—nay, all of them—will some day yield to better understanding of the phenomena in question. In the meantime, we believe that they all merit attention, that each one of them offers something worthwhile to think about, to discuss, and to try to apply to still newer problems.

Notes

1. And I might add at this point that in my view CS should continue to resist endeavoring to “exhaust the information a speaker needs in order to use [an] element... in accordance with the conventional expectations of the speech community.” That would involve building those expectations into the linguistic system, an enterprise that would be doomed to failure, since every little speech community—from skateboarders to stockbrokers—has different conventions, different expectations, different things to talk about, without necessarily having to have different grammars to talk about them. What’s in Focus, or more Relevant, or sufficiently Differentiated to the boader might not be to the broker—or, for that matter, to the broker when he’s hanging with his boarding buddies.

2. To respond: My proposal would not “further restrict” the range of CS responsibility so much as maintain its most conservative limits, i.e., the distribution of forms in text.

3. CS or CS-influenced analyses have, however, made use of degree of interlock, or information load, or opposition of substance; cf. García (1983), Davis (1995, 2002a). And Tobin (e.g., 1995) does make use of markedness.

References


This paper argues that William Diver’s signal-meaning pair is Saussure’s *signe linguistique* in all basic respects, and that Diver’s innovation of a grammatical system is the functional equivalent of Saussure’s *langue*. Thus Columbia School linguistics rests squarely on a Saussurean foundation. In the course of making this case, this paper proposes a resolution of the apparent contradiction between Saussure’s definition of the linguistic sign in terms of substance – the union of concept and acoustic image – and his dictum that “*la langue* is a form not a substance”.

1. Introduction

Conventional wisdom has long been that William Diver’s (1921–1995) early thought was heavily influenced by Ferdinand de Saussure (1857–1913), and that Columbia School linguistics firmly rests on a Saussurean foundation. Diver himself freely acknowledged this intellectual heritage in his teaching in the late 1960s, and Columbia School’s Saussurean roots are well in evidence in Diver (1974). But over the years Diver increasingly emphasized his divergence from Saussure, so that in Diver (1995) his expressed debt to the linguist credited with founding modern linguistics is minimal. Davis (2004) skillfully argues this revisionist position in a paper entitled “The linguistics of William Diver and the linguistics of Ferdinand de Saussure”. Its most provocative claims are (1) that Saussure’s *signe linguistique* and Diver’s signal-meaning pair are fundamentally different constructs, and as a consequence Columbia School (CS) linguistics is not sign-based; and (2) that Diver came to reject Saussure’s notion of *langue*.

This paper is a defense of the conventional wisdom. I will be arguing that: (a) CS’s signal-meaning pair is Saussure’s *signe* in every major respect; and that (b) Diver revised Saussure’s notion of *langue* but did not reject it. For some readers, this may seem a sterile historical exercise. But even for them I do have one novel
I believe I can now make sense of Saussure's puzzling dictum that "la langue is a form not a substance".

2. **Langage, langue and parole**

*Langue* is a foundational concept for Saussure. After a lifetime of studying language, Saussure had come to realize that the field lacked a clear theoretical foundation. *Language*, in the layman's sense, what Saussure called *langage*, is a heterogeneous grab bag of phenomena – part acoustical, part physiological, part psychological, part social – a mixture that cannot be studied as a single thing. Saussure saw the need to isolate some part of this mass as the unique domain of linguistics. The part he selected was the semiotic. This part, though psychological in nature, cannot be studied by psychology because at the level of the individual there are no units; every utterance a speaker makes is conceptually and phonetically different. The purely psychological and individual aspect of language Saussure called *parole*, speech. Linguistic units, by contrast, emerge from a social perspective because it takes a community of speakers to establish and maintain them. The two linguistic utterances [mæn] and [mæ:n] are the same word not because they are phonetically identical – they are not – but because we treat the difference between them as insignificant. On the other hand, the two utterances [mæn] and [men] count as different words despite their phonetic similarity because we treat the difference between them as significant. This social aspect of language Saussure dubbed *langue*. This, he said, was the proper domain of linguistics and the linguistic object to which all other aspects of language must be related. Roy Harris, in his 1986 English version of the *Cours*, translates *langue* as 'linguistic structure', which strikes me as felicitous. Certainly most linguists would concur that linguistic structure is their object of study (Otheguy 2002: 378).

3. **Criticisms of langue**

There have been two major criticisms of *langue*. For Saussure, *la langue* was something existing only imperfectly in any single individual; it existed in its complete form as a kind of collective knowledge of the community as a whole. This feature opens Saussure to the charge of idealism. Harris (1990) pushes this criticism forcefully, arguing that while idealization is legitimate in certain circumstances, Saussure used it to create an initial object of study that has no definable material existence. The result is a putatively empirical science founded to explain and elucidate a fiction. The second criticism is that the notion of a single *langue* shared by
Chapter 1. Columbia School and Saussure’s *langue*

An entire community is at odds with the facts of linguistic variation. When we get down to it, no two people speak exactly the same.

The validity of these criticisms hinges on exactly what *langue* is taken to refer to. If Saussure intended it to correspond to (the linguistic structure of) a *language* in the layman’s sense, say, English, French, or Latin, then the criticisms are devastating indeed. Yet it is hard to believe that Saussure was so naïve in the Introduction of the *Cours* when he shows himself to be so knowledgeable in Part Four: Geographical Linguistics. In a section entitled “Dialects have no natural boundaries” Saussure points out that isoglosses do not bundle and that as a consequence dialects do not exist as homogeneous linguistic objects with discrete geographical boundaries; what exist are only the individual dialect features. (The translations of all cited passages are those of Roy Harris (H). The corresponding French original is identified by a page reference to the Payot (P) edition of the *Cours*.)

[Dialects] are envisaged as clearly defined linguistic types, determinate in all respects and occupying areas on a map which are contiguous and distinct.

But natural dialect changes give a quite different result. As soon as linguistics began to study each individual feature and establish its geographical distribution, the old notion of a dialect had to be replaced by a new one, which can be defined as follows: there are no natural dialects, but only natural dialect features. Or – which comes to the same thing – there are as many dialects as there are places. (H 200, P 275–6)

In Saussure’s next section, “Languages have no natural boundaries”, he says the same holds for languages as for dialects, but on a larger scale:

[I]t is no more feasible to determine boundaries separating related languages than to determine dialect boundaries. The extent of the area makes no difference. Just as one cannot say where High German ends and Low German begins, so also it is impossible to establish a line of demarcation between German and Dutch, or between French and Italian. Taking points far enough apart, it is possible to say with certainty ‘French is spoken here; Italian is spoken there’. But in the intervening regions, the distinction becomes blurred. The notion of smaller, compact interme-
Immediate zones acting as linguistic areas of transition (for example, Provençal as a halfway house between French and Italian) is not realistic either. In any case, it is impossible to imagine in any shape or form a precise linguistic boundary dividing an area covered throughout by evenly differentiated dialects. Language boundaries, just like dialect boundaries, get lost in these transitions. Just as dialects are only arbitrary subdivisions of the entire surface covered by a language, so the boundaries held to separate two languages can only be conventional ones. (H 202, P 278–9)

These two passages show that Saussure was well aware that speech varies continuously; it is thus unlikely that he equated *la langue* with the layperson’s notion of either a language (e.g., French, Italian, German) or a dialect (e.g., Provençal). So what did Saussure intend by *la langue* if not a dialect and not a language? I believe the answer lies in the last line of the first passage above: “there are as many dialects as there are places.” In other words, *la langue* was a linguistic object shared more or less by speakers in a single sedentary, rural community, a village of the type Gilliéron canvassed in compiling his *Linguistic Atlas of France*.

4. **Features of Saussure’s langue**

The concept of *la langue* has several related features which must be unpacked and distinguished in order to determine the extent to which Diver and CS linguists in general subscribe to it. Listed below are statements in the *Cours* describing *la langue*. The translations are part mine and part Harris’; but I am retaining for the moment the term *la langue* in place of Harris’ translation so as not to favor my case at the outset.

1. While the totality of language (*le langage*) is heterogeneous, there exists something within language that is homogeneous, *la langue*. (H 14, P 32)
2. *La langue* is a system of signs in which the one essential is the union of sense (*sens*) and sound pattern (*image acoustique*). (H 14, P 32)
3. *La langue* is structured and self-contained, and is a system of classification. (H 10, P 25)
4. *La langue* (being homogeneous) is the (proper) object of study for linguistics. (H 10, P 25)
5. *La langue* is a psychological object, residing, albeit imperfectly, in the brains of individuals. (H 12–14, P 29–32)
6. Members of a linguistic community share the same signs, more or less. (H 13, P 30)
7. *La langue* is thus also a social object. (H 13, P 30)
8. *La langue* is a form and not a substance. (H 120, P 169)
In my judgment, Diver rejected feature 4; revised features 2 and 8; and maintained the rest. With 8, however, there is a complication. Interpreted out of context, it doesn’t, to my mind, accurately represent Saussure’s thought. Later, I will explicate Saussure’s position, and also Diver’s revision of this concept. The first revision to be discussed involves the conception of *langue* as a single linguistic system, feature 2.

5. **Grammatical systems**

During the 1960s Diver began to posit a new kind of grammatical construct in his analyses of English, Latin, and Greek. These constructs, which he called grammatical systems, were innovative in that their structure, though formally statable, was essentially semantic. Each system was defined by its *semantic substance*: a conceptually homogeneous area such as Time, Number of Entities, Probability of Occurrence, Degree of Control (of an Event). The semantic substance of each such system was then subdivided by the linguistic *meanings* signaled by specific forms in the language. Here, in reduced form, is an example.

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Signal</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAST</td>
<td>-ed</td>
<td>arrived</td>
</tr>
<tr>
<td>NON-PAST</td>
<td>-Ø</td>
<td>arriveØ</td>
</tr>
</tbody>
</table>

**Figure 1.** The English Time System

The semantic substance of Time is subdivided by the meanings *PAST* and *NON-PAST*, each signaled by specific verb morphology in English (counting a zero signal as morphology). Reid (1991) provides several more examples of grammatical systems adapted from Diver’s early analyses of English.

<table>
<thead>
<tr>
<th>Number of Entities</th>
<th>Signal</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONE</td>
<td>-Ø</td>
<td>chickenØ</td>
</tr>
<tr>
<td>MORE THAN ONE</td>
<td>-s</td>
<td>chickens</td>
</tr>
</tbody>
</table>

**Figure 2.** The English Entity Number System
The -s of *chickens* in ‘he fed the chickens’ signals more than one and its absence on *chickenØ* in ‘he fed the chickenØ’ signals one.

<table>
<thead>
<tr>
<th>Differentiation</th>
<th>REQUIRED</th>
<th>SUFFICIENT</th>
<th>the</th>
<th>the chicken</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>INSUFFICIENT</td>
<td>a</td>
<td>a chicken</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ø</td>
<td>Ø chicken</td>
</tr>
</tbody>
</table>

**Figure 3.** The English Differentiation System (singular only)

The semantic substance of Differentiation concerns the interpretive problem for the hearer of identifying the intended referent of a noun. *The* signals that finding its referent is a problem (Differentiation is required); and, moreover, that the communicative context (either linguistic or extra-linguistic) provides sufficient information to make a correct identification. *A* signals that the noun poses a problem but the context does not provide sufficient information (Differentiation required, insufficient). The zero signal indicates that no specific differentiating information is required to interpret the noun (Differentiation not required). This can be for a variety of reasons: the referent is completely obvious (‘he went Ø home’); identifying the referent is irrelevant to the communication (‘he went to Ø jail; he came by Ø train’); the noun only has one referent (‘The United States invaded Ø Iraq’); the speaker is not differentiating among various possible instances (‘Ø Water is a compound; Ø Sincerity is a virtue’). Note that the meanings in this system all address an interpretive problem for the hearer rather than being semantic components of the message itself. Huffman (1997) has dubbed this innovation *instrumental meaning*.

In the systems shown above, the signals are all discrete morphemes. Diver also posited features of word order as signals of grammatical meanings. For example, in his Kind of Differentiation System, the position of the adjective (the differentiator) before or after the noun (the differentiated) signals meanings that specify the semantic function of the adjective in the message.
Figure 4. The English Kind of Differentiation System

In ‘the boys left the windowless house’ the position of windowless (the differentiator) preceding house (the differentiated) signals that windowless is functioning to differentiate the referent of the word house from other possible referents of that word. In ‘the boys left the house windowless’ the position of windowless (the differentiator) after house (the differentiated) signals that windowless is functioning to differentiate house from itself (presumably in some other state or at some other time). Note here that an important feature of literal sentence ‘meaning’, namely that the boys broke the windows, is not actually encoded linguistically. Rather, it is an inference from being told that windowless differentiates the house from itself (implying a change of state) and our familiarity with the proclivities of boys.

In positing meanings that address a processing problem and in treating features of the message as inferences, Diver was integrating the needs and abilities of language users into linguistic structure. He called this intrusion of people at the ground level the human factor. As the human factor assumed greater importance in his analyses, he came to abandon the conventional view of a language as an abstract representational system of sentential meaning in favor of language as a rough and ready communicative tool that provides hearers with hints and instructions for constructing a message but that does not actually encode it.

6. Diver’s revision of Saussure’s langue

The discovery of grammatical systems led Diver to revise the Saussurean notion of langue from that of a single, highly-integrated sign system to a heterogeneous assembly of structurally independent sub-systems, along with a largely unstructured collection of lexical items. Each such sub-system could now be analyzed, stated and tested independent of the others. The theoretical significance of this advance is that it allows the linguist to relinquish the traditional notion of a language shared by a community of speakers, without concomitantly depriving linguistics of an object of study. Something linguistic still remains that is significantly social; namely,
individual grammatical systems. If one could look at the linguistic inventories of individual speakers, each inventory is unique, even those of rural peasants in nineteenth century French villages. In this sense, each person speaks his own language. Yet when one examines the structural components of these unique idiolects, each individual component is widely shared. And it is these (shared) individual structural components that linguists study, not any single idiolect in its totality.

For example, each of the English grammatical systems listed above is shared by a group of speakers; and this makes it possible to pool data from many sources in analyzing each of them. In doing this, it is not necessary to claim that the group sharing one system shares the others as well. In other words, it is not necessary to posit (or assume) the existence of a single community of speakers who share the same collection of grammatical systems and lexical items as a totality (i.e., speak the same language) in order to do linguistic analysis. It is perfectly possible for a group of speakers to share, say, the Differentiation system without sharing the Entity Number and Time systems.

7. Columbia School’s Saussurean heritage

Let us now return to the question of whether CS subscribes to Saussure’s notion of langue. Diver’s discovery of grammatical systems created a linguistics no longer dependent on the mystical notion of a language hovering just out of reach above its speakers, yet a linguistics that is, I believe, still Saussurean. In brief, I see a CS grammatical system as the functional equivalent of Saussure’s langue. My argument, in Saussurean style, is in terms of a parallelism of value and an identity of substance. As for value, a CS grammatical system would count as only a component of Saussure’s langue; but just as la langue was the largest linguistic object Saussure envisioned as an object of study, so a grammatical system is the largest linguistic construct the CS linguist posits as a testable hypothesis. As for substance, the characteristics of a CS grammatical system closely parallel those Saussure predicated of langue. To make this case, I will repeat below the eight separate features of la langue according to the Cours, but this time substitute ‘linguistic structure’ (Harris’ English gloss) for la langue in features 1 and 2, and subsequently substitute Saussure’s own definition of la langue as ‘a sign system’ in features 3–8.

1. While the totality of language is heterogeneous, there exists something within language that is homogeneous, linguistic structure.
2. Linguistic structure is a system composed of signs in which the one essential is the union of sense and sound pattern.
3. A sign system is structured and self-contained, and is a system of classification.
4. A sign system (being homogeneous) is the (proper) object of study for linguistics.

5. A sign system is a psychological object, residing, albeit imperfectly, in the brains of individuals.

6. Members of a linguistic community share the same sign system, more or less.

7. A sign system is thus also a social object.

8. A sign system is a form, not a substance.

What, now, needs to be changed on this list to make it apply to Columbia School? Feature 1 reflects CS thinking as written; but Feature 2 calls for revision. CS posits discrete sign systems defined and distinguished by the different semantic substances their meanings categorize. To incorporate this advance into the list, 2 must be revised to read:

2a. Linguistic structure consists of discrete grammatical systems composed of signs in which the one essential is the union of sense and sound pattern.

The significant change is the shift of the word *system* from singular to plural. This change reflects Diver’s discovery of discrete grammatical systems, which made it possible to discard Saussure’s notion of a person’s idiolect as a single system. This revision purges the notion of a shared sign system of idealism and renders it compatible with the continuum of linguistic variation both within and between speech communities. The phrases ‘albeit imperfectly’ and ‘more or less’ can now be dropped from 5 and 6, since those were Saussure’s attempt to accommodate linguistic variation while still maintaining *la langue* as encompassing the entire shared linguistic repertoire of a speech community.

8. *Langue* as the object of study

Feature 4, regarding *langue* as the proper object of study of linguistics, is where Diver came to believe he departed from Saussure most sharply. As presented in the *Cours, la langue* is an abstract entity assumed to exist in advance of analysis. But it is unfortunate to erect the foundations of an empirical science upon what appears to be an *a priori* assumption. Diver had long been concerned about the epistemological foundations of linguistics, and arguably his most original theoretical innovation was to invent a linguistics free of *a priori* abstractions and to make actual language use the ultimate object of linguistic study, its *explicandum* (see Huffman, this volume). Diver’s austere phrase was “the acoustic asymmetry of

To understand his intent here, we must recall the times. It was the late 1960’s, the heyday of the Chomskyan revolution in which Chomsky had redefined linguistics as a branch of cognitive psychology. The most fundamental problem for this newly conceived discipline was to account for the grammaticality of sentences. In that enterprise, the raw data were to be grammaticality judgments about (fabricated) sentences, speakers’ intuitive sense of sentence structure, and their sense of the structural relations between sentences.

The sentence structure Chomsky had in mind here was essentially that found in any nineteenth century grammar book. But Diver, trained as an Indo-Europeanist, knew Latin and Greek well and was intimately familiar with the shortcomings of even the most comprehensive and scholarly traditional grammars of the classical languages. As he saw it, these grammars consist of a small component, the grammar proper, which presents a version of sentence structure based on an introspective, quasi-logical analysis of a complete thought. But this component, while internally consistent and aesthetically appealing, is massively contradicted by every known text. The grammars also contain a much larger component, the appendix, which catalogues all the empirical failings of the grammar proper. For example, the grammar proper of Latin contains the rules: ‘The accusative is the case of the direct object’; and, ‘the dative is the case of the indirect object’. The appendix then lists the large number of verbs that take their direct objects in the dative or some other case and the many uses of the dative for things other than the indirect object. Finally, there is a historical component, which attempts to explain the discrepancy between the grammar proper and the appendix as due to historical change from a time when the grammar fit the language better.

Diver found all this to be an incoherent mishmash, a remnant of Greek philosophy and pre-scientific thinking that had been perpetuated in our educational system through the teaching of Latin, and that has continued to shape our thinking about both language and grammar. Yet, traditional grammar’s patina of Platonic idealism has blinded even modern linguists to its empirical failure. For Diver, then, Chomsky’s call to make sentence structure and grammaticality linguists’ object of study was to found the discipline on a faulty analysis of language. Diver’s intent in making the sound of speech its object of study was to bypass that grammatical tradition and make linguistics ultimately responsible to pre-theoretically observable phenomena, thus putting it on a par with the natural sciences. The corollary is that linguistic structure (in the form of grammatical systems) is bumped up the theoretical ladder to become (a major component in) the explanation of speech. The sound of a stretch of speech is determined by the particular sequence of signals the speaker has chosen which, in turn, is determined by their
systemic meanings and the contribution of those meanings to the communication of a particular message. In summary, for Diver, linguistic structure is no longer the explicandum of linguistics, as it was for Saussure (and Chomsky), but the explicans. To incorporate this innovation into our list, feature 4 must be discarded and replaced by:

4a. These systems are hypotheses that are tested against instances of actual language use, which comprise the ultimate phenomena of linguistics.

9. ‘La langue is a form, not a substance’

I have saved for last the most intractable item, feature 8: “la langue is a form, not a substance”. Here too Diver saw himself departing from Saussure. But before addressing the question of a possible departure from Saussure, a more pressing issue must be addressed, that of internal contradiction within the Cours itself. There la langue is repeatedly described as lacking substance:

The language itself (la langue) can be nothing other than a system of pure values. (H 110, P155)

In language itself (la langue) there are only differences. (H 118, P 166)

the language itself (la langue) is a form, not a substance. (H 120, P 169)

Yet linguistic signs, the units that comprise la langue, are repeatedly described as having substance.

In our terminology, a sign is the combination of a concept and a sound pattern. (H 67, P 99)

A linguistic sign is not a link between a thing and a name, but between a concept and a sound pattern. (H 66, P 98)

A language system (langue) is a system of signs in which the one essential is the union of sense and sound pattern. (H 14, P 32)

[A language system] can be localized in that particular section of the speech circuit where sound patterns are associated with concepts. (H 14, P 31)

It's hard not to see a contradiction here. Yet it's also hard to believe that a contradiction of such magnitude could have been inadvertent. Adding to the puzzle, this apparent contradiction can be found in Saussure's discussion of the sign itself. As seen above, he defines the sign as the combination of a concept and a sound pat-
tern, so signs clearly have both conceptual and phonological substance. Yet later in the text he seems to remove the substance.

…the concepts in question are purely differential. That is to say, they are concepts defined not positively, in terms of their content, but negatively by contrast with other items in the same system. What characterizes each most exactly is being whatever the others are not. (H 115, P 162)

Linguistic signals are not in essence phonetic. They are not physical in any way. They are constituted solely by differences which distinguish one such sound pattern from another. (H 117, P 164)

What can we make of this? To understand Saussure’s thought we must stop treating the Cours as an unordered set of axioms and definitions and read it as a document that is trying to communicate a complex set of interrelated ideas that must nevertheless be presented sequentially. To begin, let us take a look at how the Cours is structured. It is divided into six parts, each of which has from two to eight chapters.

- Introduction
- Part One: General Principles
- Part Two: Synchronic Linguistics
- Part Three: Diachronic Linguistics
- Part Four: Geographical Linguistics
- Part Five: Conclusion

The Introduction addresses the foundational question of the proper object of study of linguistics, and General Principles examines the nature of its basic unit. Here we find definitions of langue, parole and signe. Part Two, Synchronic Linguistics, then turns to the methodological and analytical issues that arise in doing actual linguistic analysis. How do we determine, describe and define the signs in a language? The statements that define signs in terms of substance occur in the Introduction and Part One, while the statements that seem to deny their substance all occur in Part Two.

This shift, I argue, is not a contradiction, but a response to different theoretical questions. Saussure’s substantive definition of a sign as the combination of a concept and a sound pattern answers the question “What is a sign?” It is a definition that states the essential features something must have in order to count as a sign. For the practicing linguist, however, the critical question is “How should an individual sign be described?” To this Saussure answers “…the concepts…are…defined negatively by contrast with other items in the same system.” This idea is paraphrased in multiple ways in the chapter on Linguistic Value in Part Two.

In a given language, all the words which express neighbouring ideas help define one another’s meaning. (H 114, P 160)
To summarize, we are concerned with two different kinds of definitions here, definitions designed to answer different questions concerning identity. The first question is one of collective identity: What makes something a sign? In other words, what distinguishes signs from non-signs? To this, Saussure answers in terms of substance: the combination of some concept and some sound pattern. The second question is one of individual identity: What makes a sign the particular sign it is? In other words, how can one sign be distinguished from another sign? To this, he answers in terms of value: its fixed relation to other signs.

10. The non-discreteness of thought and sound

One might well wonder why Saussure did not answer the question of particular identity in terms of substance as well. Why did he not say: “the identity of a sign resides in the particular concept and the particular sound pattern that are combined”? Here we must look at Saussure’s view of thought and sound, each considered in itself.

Psychologically, setting aside its expression in words, our thought is simply a vague, shapeless mass… In itself, thought is like a swirling cloud, where no shape is intrinsically determinate. No ideas are established in advance, and nothing is distinct, before the introduction of linguistic structure (langue).

But do sounds, which lie outside this nebulous world of thought, in themselves constitute entities established in advance? No more than ideas do. The substance of sound is no more fixed or rigid than that of thought. (H 110, P 155)

Neither concepts nor sounds can be precisely described, says Saussure, because they are too amorphous. How, then, do his signs acquire substance if they are defined in purely relational terms? The answer is a striking image that follows the passage above.

So we can envisage the linguistic phenomenon in its entirety – the language, that is – as a series of adjoining subdivisions simultaneously imprinted both on the plane of vague amorphous thought, and on the equally featureless plane of sound. This can be represented very approximately as in the following sketch. (H 110, P 155–6.)
At the risk of muddying the waters, I will offer my own analogy here. The state of Colorado is a perfect rectangle whose sides correspond to no feature of the natural landscape like a river or mountain range. It can thus be easily described in terms of the latitude and longitude values of its four sides. Alternatively, it could be described in purely negative terms as the territory bounded by Kansas, Nebraska, Utah, New Mexico, Oklahoma, and Wyoming, as shown in Figure 5 below. Both these descriptions function to differentiate Colorado from the other American states. Note that neither says a word about the unique topographical and geological features of the state such as the Rocky Mountains. Nevertheless, it would be wrongheaded to say that the political entity the state of Colorado lacks them. Colorado has geological substance because its political boundaries are inscribed upon the continent of North America and they circumscribe a unique and definite terrain.

In a similar fashion, individual linguistic signs have substance because their relationally defined boundaries are inscribed on planes of conceptual and phonological substance. The continued presence of substance in signs is repeatedly acknowledged in the same chapter that also says *langue* is a form and not a substance. I have put in bold type the phrases that show this:

Furthermore, the notion of value, thus defined, shows us that it is a great mistake to consider a sign as **nothing more than the combination of a certain sound and a certain concept**. (H 112, P 157)

The content of a word is determined in the final analysis not by what it contains but by what exists outside it. As an element in a system, the word has **not only a meaning** (*signification*) but also – above all – a value. (H 114, P 160)

A linguistic system is a series of phonetic differences matched with a series of conceptual differences. But this matching of a certain number of auditory signals and a **similar number of items carved out from the mass of thought** gives rise to a system of values. (H 118, P 166)
To be sure, the three quotes all affirm the primacy of value in the linguistic sign; but each still acknowledges (in the boldface phrases) the presence of substance.

Figure 5. U.S. Western states

11. Saussure’s anti-nomenclaturism

If my explication of the apparent contradiction in the Cours is correct, one can fairly ask why Saussure allowed it to arise. Why did he opt for the extreme statement “la langue is a form not a substance”? My answer is pedagogical hyperbole.\textsuperscript{13} The text of the Cours was reconstructed by Saussure’s colleagues from classroom notes, and it still retains features of its source. In many places it reads like a tract, combating conventional ideas about language as one might do in the classroom setting. Chief among these is nomenclaturism. Saussure attacks nomenclaturism in the very first sentence of Part One, General Principles.

For some people a language, reduced to its essentials, is a nomenclature: a list of terms corresponding to a list of things…This conception is open to a number of objections. It assumes that ideas already exist independently of words. (H 65, P 97)
Two paragraphs later he restates his anti-nomenclaturist position.

A linguistic sign is not a link between a thing and a name, but between a concept and a sound pattern. (H 66, P 98)

How does Saussure’s anti-nomenclaturism lead to hyperbole? Recall, for him both conceptual and phonetic substance are vague and shapeless, resistant to substantive description. Thus, defining a sign, even partially, in terms of its substance would be conceding that conceptual and phonetic substance are not so amorphous after all, that each can be described in itself. This, for Saussure, would be a concession to nomenclaturism. Saussure returns to this point again and again. I have put in bold type the phrases that evince his preoccupation with repudiating nomenclaturism:

Value, in its conceptual aspect, is doubtless part of meaning. It is by no means easy, indeed, to draw the distinction in view of this interconnection. Yet it must be drawn, if a language is not to be reduced to a mere nomenclature. (H112, P158)

These categories are difficult for a Frenchman, because his language does not recognize them. If they were predetermined categories, there would be no such difficulty. In all these cases what we find, instead of ideas given in advance, are values emanating from a linguistic system. (H 115, P 162).

Whether we take the signification (signifié) or the signal (signifiant), the language includes neither ideas nor sounds existing prior to the linguistic system, but only conceptual and phonetic differences arising out of that system. (H 118, P 166)

Saussure, it seems, saw no middle ground between nomenclaturism and purely relational definition. To summarize, I am arguing that Saussure’s famous dictum “la langue is a form, not a substance” is misleading, because it implies that Saussure’s signe linguistique had no substance, which is not the case. A more accurate version of Saussure’s thought – and feature 8 – is the following:

8a. Linguistic signs must be described in purely relational terms, not in terms of their conceptual and phonological content.

Having come to a better (and non-contradictory) understanding of Saussure’s thought in the form of 8a, we can now address the question of Diver’s possible departure from Saussure on this feature. Both Saussure’s langue and Diver’s grammatical systems have conceptual content, so Diver did not actually restore substance to a Saussurean system of pure value.14 Where Diver differs from Saussure is that he found it both feasible and necessary to describe individual signs in terms of both their substance and their value. Recall Diver’s grammatical systems displayed above. Their semantic substance is defined in purely conceptual terms (e.g., Time,
Entity Number), and their meanings have both substantive and relational components: past, non-past; required, not required; sufficient, insufficient. Thought isn’t so vague and shapeless after all; it can, to some extent, be described in positive terms. So while both Diver’s and Saussure’s sign systems have substance, Diver’s do so explicitly. Thus, feature 8a should be rewritten as:

8b. Linguistic signs must be described in terms of both their (conceptual and phonological) substance and their relation to the other signs in the system.

12. Columbia School linguistic structure

We have worked our way through eight features of Saussure’s *langue* with an eye to their presence in Diver’s work and thought. Three have been changed so as to reflect divergences from Saussure. Features 2a, and 8b are revised versions of 2 and 8a appearing on the earlier lists, while 4a is an out-and-out replacement of 4. The remaining features of Saussure’s *langue* hold, I believe, for a CS grammatical system, the functional equivalent of Saussure’s *langue*. On the list below I have expanded and revised their wording so as to articulate the precise way in which a CS system exemplifies each feature.

1. Underlying heterogeneous language phenomena, there exists something homogeneous, linguistic structure.

2a. Such structure takes the form of discrete grammatical systems composed of signs, defined as the union of a concept (a CS meaning) and a signal.

3. Each grammatical system is a self-contained principle of semantic classification consisting of meanings that exhaustively categorize a particular semantic substance.

4a. These systems are hypotheses that are tested against instances of actual language use, which comprises the ultimate phenomena of linguistics.

5. Each such system has psychological reality in that it appears to guide and explain the linguistic choices of an individual speaker.

6. Many speakers appear to share the same system, allowing us to pool data from many sources.

7. Because we often pool data from many sources, we effectively operate on the assumption that the knowledge of any particular grammatical system is social.

8b. Linguistic signs are described in terms of both their (conceptual and phonological) substance and their relation to the other signs in the system.
13. Evaluation

What now is the verdict? With three out of eight features of Saussure’s *langue* altered, does this count as acceptance? Acceptance with revision? Or rejection? Before answering, permit me, first, a brief digression and then a sketch of my own picture of Diver’s achievements. Columbia School practitioners are fond of pointing out that Diver did actual linguistic analysis whereas Saussure’s *Cours* is speculative and programmatic. This is an important distinction. In a discipline that aspires to be an empirical science, ideas, by themselves, have little value; they must bear fruit in the form of successful analysis. So if Diver and Saussure had been contemporaries, Diver would get the prize and Saussure might have been reduced to a footnote. But Diver began his innovative work fifty years after the appearance of the *Cours* and thus had the advantage of picking up where Saussure left off; he did not have to invent his conceptual framework from scratch. Naturally enough, definitional unclarities of key Saussurean constructs had to be resolved and many features would evolve in the course of putting the framework into practice. For these resolutions and development Diver gets full credit. But they do not lessen Diver’s intellectual debt to Saussure since the opportunity to develop and refine Saussure’s speculative framework was due entirely to an accident of birth.¹⁵

The ideas in the *Cours*, while innovative and provocative, are too programmatic, inconsistent and contradictory to constitute a feasible research program. Diver’s great achievement was to make of them a coherent framework for successful linguistic analysis by adding crucial ideas and insights truly his own. To summarize:

- Diver developed a sign-based conception of grammatical structure (i.e., individual grammatical systems) that, while fully Saussurean in spirit, is scarcely hinted at in the *Cours*;
- Diver’s word-order signals provided a sign-based way to handle asymmetries of word order without invoking the sentence-based syntax that Saussure had rejected;
- Diver reconceptualized the epistemology of linguistics by founding it on directly observable phenomena rather than the *a priori* abstractions of grammaticality and sentence structure;¹⁶
- Diver replaced what Harris (1990) calls Saussure’s *telementational* model of language with an *inferential* model, one in which language no longer represents thought; instead, it merely contributes to the communication of thought;¹⁷
- To make this new model work Diver introduced the distinction between *meaning* and *message* and introduced *instrumental* meaning; without these innovations, Saussure’s *signe linguistique* with a fixed and stable *signifié* would be empirically untenable;
This new analytical framework wiped out a number of distinctions that have traditionally structured linguistic analysis: sentence and discourse; semantics and pragmatics; denotation and connotation; sentence meaning and utterance meaning;

In making speakers’ choice of signs the (theory-defined) problem to be solved rather than attempting to offer a theory of communication, Diver found a way to recognize the pervasive role of human inference in the communicative process (i.e., the human factor) without having to provide a formal theoretical model of its role;

Diver developed a quantitative methodology for testing grammatical hypotheses that brought his linguistics into line with the natural sciences.

These innovations stand on a par with those of Saussure. But – and here I return to the issue at hand – he threw out surprisingly little. Consider once again Diver’s three alterations of langue. Two emerged from his successful analytical work. Diver’s early discoveries allowed linguistic structure to be reconceived as a collection of discrete sub-systems (2a) rather than as a single super-system. This, I would argue, counts as an evolution of a Saussurean idea, not a rejection. Recall, Saussure’s main objective was to provide linguistics with a homogeneous object of study, not to legitimate the folk notion of a language. A CS grammatical system gives Saussure what he was looking for without metaphysics.

Similarly, defining signs in terms of both substance and value (8b) counts as revision not rejection because, to recall again, Saussure’s signe already had substance, what Saussure called its signification. But Saussure feared that admitting substance into the actual description of individual signs would undermine the importance of value and return language to a nomenclature. Diver discovered this not to be the case; through his analytical work he found that signs could be partially described in substantive terms while still keeping value pre-eminent and nomenclature at bay. This should count as an important methodological innovation but not a repudiation of Saussure’s basic construct.  

The one feature of Saussurean thinking that Diver truly rejected is feature 4: ‘La langue (being homogeneous) is the (proper) object of study for linguistics’. For Diver, linguistics begins with speech, not sign systems, sentences, or, indeed, any version of linguistic structure. Note, however, that Diver’s claim of departing from Saussure on this point requires interpreting the phrase ‘the (proper) object of study’ as referring to the starting point of linguistic analysis. This is the interpretation Diver chose, and I have honored it here. But it is not the most charitable interpretation; and it is perhaps anachronistic as well. Saussure’s thinking was far too programmatic to be concerned with drawing the clear distinction Diver introduced some sixty years later between observations and hypotheses in CS linguistics. It is...
likely Saussure simply meant by his feature 4 that characterizing langue was to be the chief enterprise of linguistics, and that he would readily concede, if pressed, that la langue was in fact a hypothesis. After all, even a sophisticated modern biologist might well say that natural selection is his object of study, while nevertheless knowing full well that natural selection is a hypothesis not an observation. 19

Be that as it may, Diver’s explicit distinction between observations and hypotheses is an important one because it underscores the fact that linguistic structure is not observable. But this fact is not a change in the conception of a sign system itself, only a shift in its ontological status. For both Saussure and Diver, a sign system is, in the end, unobservable yet real. The two disagree only on when its reality can be definitively asserted. In Diver’s reading of Saussure, la langue exists prior to analysis, whereas for Diver its existence is truly known only after determining its individual structural features in detail.

This hard-nosed position served as the starting point for Diver’s incisive critique of the western grammatical tradition, for it applies indifferently to any and all conceptions of linguistic structure, from traditional grammar’s parts of speech to Chomsky’s Universal Grammar. All are constructs of the linguist and none can be said to exist until they have survived rigorous testing. This iconoclasm should count as one of Diver’s major contributions. But it is not, to repeat, a change in any essential characteristic of sign systems themselves. My verdict regarding Diver and Saussure’s langue, then, is acceptance with revision.

14. Conclusion

In the essay that has sparked this paper Davis (2004: 308) writes: “It would be a disservice to Diver and to the field if he were to be considered primarily as an heir to Saussure rather than as an innovator fully in his own right.” With this I agree. But it is also true that portraying Diver as an innovator does not require minimizing his intellectual heritage. We can thus recognize without slight to the originality of its founder that Columbia School’s signal-meaning pair is Saussure’s signe linguistique in all basic respects, and that Columbia School linguistics rests squarely on a Saussurean foundation.

Notes

2. Diver’s debt to Saussure became relevant when the Columbia School website was under construction and a debate arose as to whether Columbia School linguistics should be described as Saussurean. Saussure means little to contemporary linguists, it was argued, and, moreover, stressing Columbia School’s Saussurean roots did not do justice to Diver’s own originality and achievement. The issue gradually shifted, however, from the practical one of public communication to the scholarly one of Diver’s actual debt to Saussure.

3. This came to be Diver’s criticism of Saussure as well; see Huffman, this volume.

4. The informs us about the word States, which poses an interpretive problem because the word can have many different referents, but whose referents are sufficiently differentiated from other possible referents of that word in this context due in part to the word United.

5. The high structural integration of Saussure’s langue stems from the fact that each sign is defined entirely in terms of its relation to the other signs in langue, its place in the system.

6. This means that the group of speakers providing data is determined, in part, by the system under analysis. This methodological principle extends down to individuals who belong to the same “community” by every imaginable social, racial, ethnic, geographic, economic, professional and educational parameter.

7. This is a major point of disagreement with Davis (2004: 315). Davis writes “If la langue is a homogeneous, socio-psychological system, then Diver (1995) has no exact counterpart to it.”

8. Alan Huffman (personal communication) has pointed out that the largest linguistic construct for CS is a grammatical interlock. But a grammatical interlock is a composite structure consisting of several grammatical systems that happen to share the same morphology as signals. It thus does not constitute a separate hypothesis about grammatical structure since it is simply the formal configuration of constructs already hypothesized.

9. In a famous passage in Aspects of the Theory of Syntax, Chomsky (1965) makes clear his acceptance of the grammatical tradition:

   The investigation of generative grammar can profitably begin with a careful analysis of the kind of information presented in traditional grammars. Adopting this as a heuristic procedure, let us consider what a traditional grammar has to say about a simple English sentence such as the following:

   (1) sincerity may frighten the boy

   Concerning this sentence, a traditional grammar might provide information of the following sort: (Chomsky 1965: 63)

   Chomsky then parses the sentence grammatically, dividing the information into three types: (a) syntactic categories: Sentence, Verb Phrase, Verb, Noun Phrase, Noun, Determiner, Article, Modal; (b) sentential relation categories: Subject, Predicate, Object, Main Verb, Subject-Verb relation; (c) lexical categories: Count, Mass, Abstract, Common, Transitive, Progressive Aspect. Chomsky then endorses this parsing:

   It seems to me that the information presented [above] is, without question substantially correct and is essential to any account of how the language is used or acquired. The main topic I should like to consider is how information of this sort can be formally presented in a structural description, and how such structural descriptions can be generated by a system of explicit rules. (Chomsky 1965: 64, emphasis added, WR)
10. See Huffman (this volume) for a brief summary of the structure and categories of the grammar proper. For a more detailed discussion and critique see Huffman 1997, Ch. 6.

11. Diver writes: "Note that the observations, the sound waves, exist independently of any attempt on our part to come to an understanding of them; they pose a problem that we may or may not undertake to solve; we do not invent them in the process of setting up a problem" (Diver 1995: 50).

12. Reid (1991: 352–370) offers a demonstration of the way in which a hypothesized grammatical system does, in fact, explain particular features of the (pre-theoretically observable) acoustic asymmetry of speech. Reid (1991) presents a sign-based analysis of two number systems in English, one including the plural -s on nouns and the other including the 3rd person -s on verbs. His demonstration takes the following form. First he calculates the ratio of nasal and sibilant sounds in the oral version of an English text of several thousand words; this ratio serves as a particular instance of the acoustic asymmetry of speech. Next, he calculates the ratio of nasal to sibilant sounds in the same text, but this time excluding all instances of the sibilant signals in his two number systems. Finally, he subtracts one ratio from the other. This difference between the two ratios is that fraction of the overall acoustic asymmetry in the text that is accounted for by the analysis.

13. Davis (2004: 316–322) has a different way of handling the apparent contradiction in Saussure. For Davis (see p. 320), Saussure confines substance to the interior of the signifié and the signifiant. Value, by contrast, distinguishes one signe – outwardly – from another. Langue is exclusively this pattern of values, regardless of the substances within the individual signes.

14. This is a major point of disagreement with Davis (2004) and other commentators. He writes: "Saussure defined his basic unit of langue, the signe, as being devoid of substance. Saussure's signe was pure value. One of the main points of Diver (1974) is the necessity to reintroduce substance into linguistic analysis. So Diver and Saussure are not operating with identical units" (p. 308).

15. It strikes this reader that much of Davis' (2004) and Huffman's (this volume) case for the non-Saussurean Diver hinges on counting developments that are still Saussurean in spirit as departures from Saussure and then practicing a kind of double-entry bookkeeping where an innovation is first credited to Diver for having achieved it and then debited to Saussure for not having achieved it. This, to repeat, would be fair if Saussure and Diver were contemporaries working independently, but not if Diver used Saussure's conceptual framework (empirically baseless though it be) as the starting point for his own work.

16. Huffman (this volume) provides a far more extensive account of Diver's innovative epistemology for linguistics than we have offered in this paper, including his notions of explanation and orientation.

17. This constitutes a rejection of the compositional nature of sentence meaning in which the literal meaning of a sentence is determined (and hence predicted) by the semantic values of its words together with its grammatical structure. In Diver's inferential model, the aggregate of linguistic meanings signaled by an utterance underdetermine the interpretation (i.e., the message).

18. Another way to understand Diver's revision is that he eliminated the need to distinguish between the signifié and the signification of a sign. This distinction has been a puzzling and troublesome one to Saussure scholars, but one Saussure was obliged to make if he wanted signs to have conceptual substance yet also be defined in terms of pure value.
19. Huffman (this volume) is not willing to treat *langue* as even a working hypothesis, and sees it (as Diver did) as no more than an a priori assumption. “The notion of *langue* did not emerge from a body of successful analyses; and until one can say what precisely is in *langue*, the actual *signifiers* and *signifieds*, one is not in a position to assert that there exists any *langue* at all.” This seems to be an unduly harsh appraisal given the context of Saussure’s *Cours*. At the time of his lectures Saussure had spent a lifetime doing historical linguistics. The principles of the Comparative Method take as their starting point a conception of linguistic structure similar in important respects to that found in the *Cours*. If (even parts of) the reconstruction of Indo-European phonology, grammar and lexicon – the great endeavor of nineteenth century linguistics – is recognized as successful, then Saussure’s conception of *langue* is indeed rooted in successful linguistic analysis, and should at least count as what Huffman calls an expectation based on analytical experience.

References

Diver’s “Theory” (1995) is the most comprehensive and, in fact, the final statement by the founder of the Columbia School of that school’s contribution to an understanding of the essential nature of language. The unifying idea that runs through this statement is Diver’s insistence that a theory of language consist of a set of conclusions drawn from a body of individual analytical successes, that it not be a collection of *a priori* categories and speculations. Diver’s anti-apriorism opens the way to understanding the workings of language in terms of innovative and language-specific categories, and it brings the normal practice of linguistics into line with that of other natural sciences.

William Diver’s paper “Theory” (Diver 1995) is a remarkable work. It distills conclusions from decades of work within the theoretical position Diver pioneered and synthesizes the major analytical achievements of the school he himself dubbed the Columbia School. It was his last and most comprehensive summary of the state of the art in Columbia School linguistics and probably his single most important publication. In particular, it incorporates many of the impressive theoretical advances that transformed what had been known in the 1960’s and ’70’s as the Form-Content School into the more mature Columbia School of the ’80’s and ’90’s. The latter period witnessed many important new analyses such as Gorup’s (1987) and Gildin’s (1989) on Event Focus in Serbo-Croatian and French respectively, Goldberg’s (1991) on Probability in Spanish, and Davis’s (1992) on Focus and semantic structure with reference to Italian pronouns. The same period saw the refinement and even discarding of some of Diver’s earlier analyses and a backing away by Diver himself from his initial endorsement of an orthodox Saussurean position (cf. Davis 2004; Reid this volume).

The paper begins with a bit of a polemic on the notion of *theory*. The polemic fades into the background as Diver summarizes the achievements of his school, but it remains subtly and importantly present throughout. “Theory” is an
important contribution to a debate, still unresolved within linguistics, over how to conduct inquiry into the nature of language in a way consistent with the spirit and methods of modern science. Diver attempts above all to emphasize the empirical and inductive process that alone can lead the field of linguistics to the kind of analytical successes Columbia School has enjoyed. He warns against bringing to the table the questions with prefabricated answers that have led and may still lead inquiry into language to stray from the discipline properly conceived.

However, this main theme of the paper, from which its title is derived, seems to have been underappreciated even by some of those who do their own analytical work within the Columbia School paradigm and who have studied the paper. “Theory” (henceforth “Diver ’95”) is published in a volume which also contains an introduction (Contini-Morava 1995, henceforth “the Introduction”) that explains and compares several examples of “linguistic sign theory”. This carefully researched and meticulously written introductory essay provides an excellent discussion contrasting generative grammar with some of the commonalities shared by these sign-oriented schools. It will undoubtedly help those trained in generative grammar to better understand this type of non-generativist work in linguistics. Moreover, the Introduction has proved of great value to Columbia School linguists in refining our own understanding of other non-generative schools, particularly the Jakobsonian and Guillaumean schools. Thus, the Introduction performs a marvelous service for linguists of all varieties.

Less helpful, however, are the parts of the Introduction that discuss the differences between the Columbia School and the Jakobsonian and Guillaumean schools. The point of scientific method that is so central a concern in Diver’s paper receives, on my view, inadequate attention in the portrayal of Columbia School that emerges from this comparison. Moreover, foundational differences over how to define the discipline of linguistics set Columbia School apart from the Jakobsonian and Guillaumean schools no less than they do from traditional and generative grammar. Consequently, the failure to specify these points in the Introduction results in Columbia School’s being inadequately differentiated from all schools that uncritically accept, ab initio, an inventory of analytical categories inherited from tradition or otherwise not inductively derived. In failing to bring this contrast to light, the Introduction understates the substantive basis of Diver’s disagreement with nearly all other practitioners in the field of linguistics today.

What follows is an attempt to distill the essence of Diver ’95, without aiming at reductionism for its own sake. These remarks will attempt to highlight the coherence in Diver’s thinking and to show that the passages dealing with methodology in Diver ’95 are not ancillary philosophical digressions in what is otherwise an exemplary presentation of his position, but that they in fact represent the unifying
idea of the paper and are critical for appreciation of defining elements of Diver’s thinking.

By the same token, the purpose here is not to take a position on whether Diver ’95 was right or wrong in its specifics; not to express agreement or disagreement; certainly not to project a theory of my own; but only to expedite understanding of what Diver wrote and to argue that a central concern of the paper – its main point, really, be it right, wrong, or somewhere in between – has been neglected. Defining and demonstrating that concern must obviously precede any debate over the adequacy or correctness of Diver’s position.

1. Diver ’95: The Main Idea

We can begin with the title of the paper. Diver very pointedly did not entitle his paper “The Theory”, and of course not “A Theory”; rather, zero article: “Ø-Theory”. He was writing a paper, in some sense, about – theory. Here are his words, telling us what, for his purposes, theory is not and what it is:

There is of course a temptation to begin by posing the question “What is a theory?” or to demand that the term “theory” be defined before we take even the first step in the discussion. But that question and that demand rest on a certain metaphysical approach to the whole problem of the acquisition of knowledge: an approach that implies the a priori existence of certain metaphysical realities, such as theories, with an accompanying analytical task of discovering the properties of those metaphysical realities. The history of the study of language has been plagued by an insistence on bringing into the discussion a priori metaphysical realities that turn out to have nothing to do with actual languages. ... The term “theory” will here be used to refer to a summary of the general characteristics of successful solutions to individual problems. It definitely will not be used to refer to a speculative hypothesis that attempts to forecast in advance an as yet unattained solution to a problem. (Diver 1995: 45)

The theory, for us, is not the beginning of analysis but the endpoint. The theory was not assumed, much less known, in advance. Rather, it gradually dawned on those of us engaged in this kind of analysis. Analytical experience with many small problems gradually built up a body of precedents to which we have turned in addressing each new problem. But even precedents thus developed do not exercise any absolute control over analytical procedure. The theory, as the embodiment of analytical experience, is constantly open to reappraisal arising out of the exigencies of new analysis. This, then, is not a treatise on the nature of theory; this is simply a report on what our analyses are about, what ideas they have led to. That is what we call theory. (Diver 1995: 52)
The insistence on an inductive (*a posteriori*) procedure is set forth at the very beginning of the paper and relentlessly recalled throughout:

A constant theme of the paper will be the nature of the relationship between analysis and theory, and an insistence that theory be guided by analysis, rather than the other way around, no matter how unfamiliar the resulting theory may appear.

We begin with a brief summary of the most fundamental points, stressing again that these are the results of analytical procedures, conclusions reached, rather than in any way *a priori* assumptions that we made up in advance. (Diver 1995: 43)

As innocent as it may sound, in today’s linguistics this is actually quite a radical position. It is certainly the case that *theory* for other linguists is just the opposite sort of thing. One “works within a theory”; a contemporary linguist might say: “I’m working on theta-role theory”, or “I’m working in GB theory”, meaning working on fleshing out a speculative program. In philosophy of science, one finds a great variety of views on the question of what constitutes a theory; but, to my knowledge, Diver’s use of the word *theory* for a summing up of the conclusions of successful analytical experience is quite original. More on this point later.

To say that Diver’s position is radically aposterioristic, insisting as it does on approaching the language problem with a dearth of inherited gear that would make other linguists feel uncomfortably naked, does not, however, imply an anti-theoretical descriptivism. The big picture is not ignored forever, it is just postponed. Perhaps the approach is best characterized as a kind of bootstrapping. Recall Diver’s words seen earlier:

Analytical experience with many small problems gradually built up a body of precedents to which we have turned in addressing each new problem. But even precedents thus developed do not exercise any absolute control over analytical procedure.

As each apparently successful analysis is achieved, another piece of the picture is allowed to emerge. The whole picture, a conclusion as to what general statements about the nature of language one may confidently make, is not likely to take shape until a considerable portion of the road has been traveled; and even then, such conclusions, being always subject to revision as new data and analysis are developed, can never be considered irrefutably final. The tentative conclusions reached along the way can, though, exert some guidance as the process unfolds. This is an essential characteristic of scientific method.

There may seem, though, to be a contradiction lurking here. If apparent accomplishments must always be viewed with reservation, how can we actually make progress by building on them? If we refuse to accept as canon that which has been approved by the community of scholars and committed to print, then to what au-
thority can we appeal for judging subsequent work? And if, on the other hand, we proceed as though our answers were final for all time, are we not guilty of the very apriorism Diver is warning against?

The paradox can be resolved if we recognize a distinction between analytical assumptions and analytical expectations. Undertaking to solve a linguistic problem with initial assumptions about what form part or all of the solution must take, in what terms or kinds of terms the analysis will be cast, constitutes the a priori approach Diver rejects. To undertake a problem with expectations, based on experience, that we acknowledge may or may not be fulfilled this time around, in contrast, is part and parcel of the bootstrapping approach.

Here, then, is another part of the Main Idea that sounds like a banality – a commonsense, uncontroversial notion. Yet in the history of linguistics, the construction of elaborate intellectual edifices upon invalid initial assumptions has been more the rule than the exception. It is not an exaggeration to say that the ghost of prescientific thinking has largely gone unexorcised in this one discipline. Diver’s passionate opposition to linguistic analysis using categories derived from traditional sentence grammar is well known and has been extensively elaborated, by Diver himself and by others. But the same holds true for other apriorisms, some of which are more modern and definitely not part of the grammatical tradition: the Prague School insistence on binary oppositions; Saussure’s idealized langue system; the compositional view of linguistic meaning held by almost all traditional and modern linguists, and its corollary, the notion that polysemy is an observable “fact” of language. These points will receive further attention below.

2. The search for the starting-point of linguistics

An important operational consequence of the bootstrapping approach is highlighted in the following quote:

This ... is what has turned out to be the hallmark of our analytical procedures: always begin with the smallest possible problem. Get it solved, and then look around for the next smallest problem, taking advantage of the leverage obtained by the solution of the previous ones. (Diver 1995: 58)

Diver urges linguists to think small, to seek linguistic function at the lowest possible level of analysis. For example, a Columbia School analysis will not begin by seeking to identify the meaning of a sentence, a construction, or an idiom. Rather, it will use the smaller units suggested by analysis at the morphemic level to begin a search for minimal signal-meaning pairs, and only then consider larger configu-
rations in terms of what can be established at this lower level. Similarly, one will not start by asking what the large-scale structure of the phonological system of a language looks like (e.g. “Does language X have an opposition of voicing or an opposition of lenition?”), before one has undertaken to determine what the best terms are with which to characterize its individual phonemes.

Emphasis on the minimal follows in an obvious way from the bootstrapping approach, which itself implies starting from the bottom up. Non-minimal units represent multiple variables. If we are starting with the assumption that we have no prior knowledge of any of the variables, each added variable multiplies the complexity of the problem, and the problem quickly becomes unmanageable. The temptation increases to gain control by admitting unvalidated assumptions about some of the variables. Any part of an analyzed unit which is not itself also an analyzed unit must be an artifact of such an a priori assumption. By this reasoning, the unique minimalism espoused here and the rejection of a priori categories and assumptions are simply two faces of the same coin.

The minimalism and the anti-apriorism of Diver ’95 are intertwined at an even deeper level. Diver grappled for a long time with foundational issues, such as the nature of observations. He thought carefully about what linguistics ought to explain, and just how far the responsibility of the analyst extends. He realized that in order to arrive at a level that avoids unvalidated initial assumptions, a level at which everyone will agree we are confronted with phenomena that are uncategorized in any way, we have to step way back to an extremely austere position, a position so austere that most linguists would be uneasy with it at best, outright dismissive of it at worst:

Let us retreat then, to the most primitive, the most austere of all positions in respect to observation that has to do with human language. What is observable ... to the uninitiated is simply the sounds themselves, acoustic phenomena that are entirely uncategorized. ... If we occupy a position of this austerity, we are immediately putting ourselves completely outside the field of linguistics, in the opinion of most other practitioners. ... we, by committing ourselves to a study based on the sound waves themselves, necessarily begin with the study of what would normally be regarded as “mere phonetics”. Comfortable or uncomfortable as such a position may be, in taking it we pose the problem as follows: in any instance of what we informally recognize as speech, what determines the form taken by the sound waves we observe? (Diver 1995: 48)

The observations confront us with chaos, and the analytical problem is to impose order on the chaos. You might be happier if I were to say ‘find order in the chaos’ rather than impose order on it, but the truth is that there is no order there; it really is chaos. (Diver 1995: 52–53)
“Find order” is another way of saying “impose pre-existing categories”; the categories must already be known if something is to be found and recognized as an instance of that which one is looking for. To “impose order”, on the other hand, means to invent whatever categories may be necessary to establish order. This entails devising, revising, adjusting and fine-tuning hypothesized categories of analysis until a really good fit with the data is achieved.

Diver is not saying in these passages that linguistics is only phonetics; he is not saying that linguists must organize their working procedures so that they have complete knowledge of the phonology of a language before they tackle its grammatical structure. Rather, the issue is one of pinning down the motivation and the justification for what the linguist does. Diver wants to know: What are we ultimately answerable to? What is it out there that requires an explanation, that motivates a discipline of linguistics in the first place? His answer: It is not, for example, a set of preexisting conceptual categories into which linguistic phenomena must be made to fit (e.g. “What counts as a ‘subject’ in language X?”), nor an attempt to demonstrate a speculative hypothesis about the nature of linguistic structure (e.g. “Language is a set of binary oppositions”, “A language is a set of sentences whose structure can be reduced to a universal set of templates and operations”, etc.). Rather, says Diver, it is the fact that when people engage in linguistic behavior (and, as we will see shortly, it is not necessary for this to be defined a priori in any formal way), they are producing sounds, and that the sound waves they produce on these occasions have skewed, not random, shapes. This noise is not white noise, nor any other of an enormous variety of noises that might strike our eardrums. What, Diver ’95 tells us to ask, explains the particular shapes that speakers impart to these sound waves?

By the same token, Diver is not insisting we follow the descriptivist procedure of working up from the bottom, in the sense that the first step is to develop an inventory of phonemes, these then being used as building blocks for the next higher level, morphemes, and so on. Operationally, we can jump in at any point. Procedure is guided by the “minimal unit” requirement: wherever in the chain we choose to begin, we aim to identify units that cannot be further analyzed. The overall procedure of linguistic analysis is not directional. The solution to a morphemic problem might serve to inform a problem in phonology, or it might be the other way around.

Multiple sets of constructs will ultimately be posited to answer the question of what explains the shape of the sound wave of speech. Ordinary signals are made up of sound. The choice of particular signals will therefore define the shape of the overall sound wave of speech at one level, and the choice of signals is governed by the meanings associated with these signals. Thus, the discovery of signals and their associated meanings accounts for one aspect of the shape of the sound wave.
At another level, we can ask why signals are made up of the particular sounds they contain, why certain sounds are favored or disfavored altogether in speech, why some sounds prefer the beginnings of signals and some the ends, why some sounds are less likely than others to occur in combination. Answers to these questions also account for an aspect of the shape of the sound wave. At yet another level, one can ask why signals are combined in particular arrangements and not in others. When this question has been successfully answered, still another aspect of the sound wave has been explained.

Given, now, that what linguists are ultimately responsible to is uncharacterized, uncategorized raw sound, the next question that arises is: How much sound are we responsible for? Where do we draw the line separating the responsibility of linguistics from the responsibilities of other disciplines?

The first approach Diver entertains is analogous to peeling an onion. We take the entirety of the observed sound wave and proceed by peeling off layer after layer that we do not want to take responsibility for until we get to where we want to begin. Thus, one could start by peeling off background noises not produced by the human voice (e.g. trucks rumbling by). Then we might peel off variations in fundamental frequency due to differences in people's voices; then those caused by a stuffed-up nose, a sneeze; then perhaps any variations in the selection of morphemes introduced in order to fit a scansion or rhyme scheme. At some point, however, we would come to a layer where people would disagree about how much to peel off. An example of this would be prosodic features, or supra-segmental phonemes. At this point, says Diver, we have no guidance as to what to peel off. Nothing in the sound itself, which is after all only raw energy, can make this decision for us; and to simply come down somewhere arbitrarily in the midst of the chaos would be pointless. The only option would be to apply some a priori criterion, one not based on inductively-derived analysis.4

The unacceptability of such a priori criteria, then, forces us to reject the “onion” approach altogether and turn to something entirely different. Here is what he says:

The alternative to peeling off from the outside the things that obviously don’t matter is to begin from the inside. In other words, instead of beginning with a total responsibility to the entire package of the sound waves and proceeding to get rid of our responsibilities step by step, we begin with zero responsibility and look for anything in the sound waves that we can account for. As we succeed with small initial steps we gradually bring segments of the chaos under control.

We are then confronted with a straightforward analytical problem: How far can we push this procedure? The answer seems to be that at any point in the progress our responsibility extends just as far as our success in analysis. How much farther it can go remains an open question.
This procedure has the very distinct advantage that we do not begin by inadvertently setting ourselves a task so impossible of achievement that we are never able even to get well started on it. It also has the advantage that as we succeed with a few small problems what remains is cast into a clearer light. ... This, in fact, is what has turned out to be the hallmark of our analytical procedures: always begin with the smallest possible problem. Get it solved, and then look around for the next smallest problem, taking advantage of the leverage obtained by the solution of the previous ones. (Diver 1995: 58)

A clear implication of what Diver says here is that we cannot begin doing linguistics with initial idealizations of any kind. Now, of course, most researchers do make initial idealizations. In our field, the most famous idealizer was undoubtedly Saussure, whose Cours at the outset posits a highly abstract psycho-social entity called *langue*, which exists in the collective mind of a society, independent of individual acts of speech, in the same way a symphony can exist regardless of whether it is ever played by musicians. The Cours resolutely detaches *langue* from the actual sounds and muscular movements made in speech acts; these are part of what Saussure called *parole*. *Langue* is a principle of classification and an abstract system of signs, each of which consists of an abstract sense and an abstract acoustic image. The Cours tells us that if one wishes to construct a theory of language, one must choose between *langue* and *parole*; one cannot take both routes simultaneously, and the true object of study of linguistics proper is *langue*.

This Saussurean idea conflicts head-on with the view expressed in Diver ’95. First of all, it is an attempt to specify a theory of language in advance of carrying out analyses. Saussure’s theory of language, in the terms of Diver ’95, is an *a priori* speculation, not a theory. The notion of *langue* did not emerge from a body of successful analyses; and until one can say what precisely is in *langue*, the actual signifiers and signifieds, one is not in a position to assert that there exists any *langue* at all.

Secondly, Saussure’s theory represents an instance of the “onion” procedure: there is an assumption that certain things can safely be peeled off at the start and declared irrelevant to the problem. On both counts, then, Diver ’95 and the Cours are in conflict.

The Introduction to the 1995 volume states that all sign-based schools of linguistics accept the *langue*/parole distinction. But it is clear that Diver ’95 does not. Here again are Diver’s own words:

We have been led to a quite different conceptualization of the notion of “language”. It seems not to be the case (by which is meant here that no evidence has been found in support of it) that there is existing out there a “language”, in the sense of some ideal construct, which it is the job of learners to learn and analysts to analyze. The learner, rather, seems to create what may be fairly individualized
techniques of communication, more or less on the model of what can be observed in the behavior of others. This has as a consequence a lack of anything even approaching absolute uniformity from individual to individual... (Diver 1995: 45)

Recent Columbia School work is characterized by an increasing acknowledgment of the idiosyncratic nature of linguistic behavior, just as other types of human behavior are idiosyncratic, with the implication that ultimately the analyst, in any given analysis, is studying the behavior of one speaker, and in fact of one speaker at one moment in time. This is perhaps even more true in phonology and lexicon than in grammar.

Another well-known idealizer was Chomsky, whose initial idealization on the first page of Aspects was apparently intended to be much like Saussure’s. In that famous passage he says:

Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors in applying his knowledge of the language in actual performance. This seems to me to have been the position of the founders of modern general linguistics, and no cogent reason for modifying it has been offered. (Chomsky 1965: 3)

Diver explicitly distances himself from this:

Note how [our] procedure differs from that of setting up an “ideal speaker/hearer”, one who is to be representative of every and all speaker/hearers. Setting up an ideal requires a priori assumptions over which the theorist has little practical control, assumptions that necessarily ignore the reality of the observations and result in analysis that is of little practical, or even theoretical, value in the long run. (Diver 1995: 60)

Here, then, is a primary and fundamental difference between Diver and Chomsky that assuredly needs to be mentioned prominently in any comparison of the two.

The austerity of Diver’s position – the exclusion of prefabricated categories of analysis, the requirement that the categories ultimately settled upon be analyzable no further, the refusal to decide at the outset that certain facts be excluded while others are included, the rejection of idealizations whose purpose is to jump-start the analytical process – all these factors are elements of a stance that enables him to identify what is pretheoretical in the study of human linguistic behavior and ultimately allow him to ask and answer the most basic question: What is it that motivates a discipline of linguistics in the first place?
3. Columbia School distinguished from other sign-oriented schools

The preceding paragraphs have focused attention on Diver’s rejection of initial assumptions and *a priori* categories, together with his interest in seeking out minimal units of analysis and small problems, as being the overarching idea of Diver ’95. This is not to deny the importance of other aspects of his position: discoveries such as the sparse and unique structuring of linguistic encoding, the pervasive influence of well-known human behavioral traits in the structure and functioning of language, and Diver’s own fascinating and inspiring analyses of Greek, Latin and English. Nevertheless, Diver’s insistence on a radically inductive approach to human language can be viewed as his most important contribution because it is the only one that forces the whole practice of linguistics to reorient itself. It is currently without a doubt the most unassailable of Diver’s positions, the one that, if followed scrupulously, leads to his other discoveries and that most clearly and sharply differentiates Diver’s linguistics from other schools. One could anticipate that other parts of Diver’s theory would eventually be hit upon by anyone who started out from the same austere beginning. Many have indeed been hit upon to some extent – signals and meanings by all Saussure’s heirs, the human factor by Zipf, the role of inference and context by anyone who has ever tried to teach a language or program a computer to process or produce natural language.

The fact is that any one of Diver’s analyses can be overturned if a better analysis comes along. What is irrefutable, enduring and profoundly original, however, is his critique of the faulty premise of an underlying logic-based structure assumed by traditional and modern thinking about language in the Western tradition. The Introduction does point out the Columbia School rejection of traditional parts of speech because of their circular and inconsistent definitions. But what needs to be pointed out, lest the case be severely understated, is that other schools’ interest in parts of speech has its origin in an *a priori* scheme, and that Diver ’95 rejects that entire scheme.

Diver insisted that since the canon of categories represented by the grammatical tradition has in fact always been used to explain occurrences of linguistic forms, it must therefore be thought of and evaluated as a theory of language. He showed how these categories are motivated by the premise that the structure of language reflects the structure of rational thought. This premise deductively motivates organizational categories like sentence, subject, predicate, direct and indirect object, as well as conceptual categories like noun, verb, and adjective. A special relationship is held to exist between these two types of categories and linguistic forms, by means of which occurrences of specific forms are rendered predictable. Thus, for example, in Classical Greek or Latin, the subject will be a noun in the nominative case; the direct object will be a noun in the accusative case; the indirect
object will be a noun in the dative case. (See Huffman 1997, Ch. 6.) In the actual usage of these languages, though, we find that in fact nominatives can occur in the predicate, direct objects in the dative. In his writings and lectures, Diver showed in great detail how this theory, which he called the Theory of the Sentence, breaks down massively and across the board; and not just in English, but specifically in the languages for which it was originally devised, Latin and Greek. Nonetheless, its unsuccessful categories survive in one form or another in almost all schools of linguistics, including the sign-oriented schools to which Columbia School is compared in the Introduction.

Take, for example, an analysis in a Jakobsonian framework. The parts of speech play a crucial and fundamental role in Waugh’s (1976) analysis of adjective-noun word order in French, a work which is cited in the Introduction. Waugh’s hypothesis is cast in terms of the traditional notion of modification, and of an intersection between the sets given by the part-of-speech adjective and the part-of-speech noun.

What is important here is that modification itself presupposes nothing more than the presence of the part-of-speech adjective and part-of-speech noun in the given syntagmatic context. Modification possibilities, then, and the semantic nature of modification itself are established with respect to parts of speech, not with respect to given lexical items. This is very important when we define the feature for preposition. (Waugh 1976: 88)

Waugh’s analysis assumes that we know in advance which of the two words under analysis is the adjective and which is the noun. Can this in fact be presupposed? She discusses, for example, the difference between *furieux menteur* ‘compulsive liar’ and *menteur furieux* ‘angry liar’. This assumes we know in advance that *menteur* is a noun and *furieux* an adjective. But it overlooks the reverse possibility, a distinct possibility in French: that *furieux menteur* means ‘lying maniac’, with *furieux* taken as noun and *menteur* as adjective. No recognition is accorded to the fact that this part of the process is based on ad hoc, local inference, not on an underlying universal categorization. Waugh overlooks the fact that the very ordering she is studying is part of what feeds into the ad hoc decision as to what is to be taken as a noun and what as an adjective.

These same categories are absolutely fundamental to other non-generative schools which often invite comparison with Columbia School, for example, Cognitive Grammar.

Nouns and verbs are the two most fundamental grammatical categories...they are pivotal to the description of every natural language. The respective counterparts of nouns and verbs at the level of sentence structure are nominals and finite clauses, whose universality and grammatical significance are also beyond dispute. (Langacker 1991: 51)
If *subject* is defined schematically as *relational figure*, it is both coherent and quite reasonable to posit subjects even for a language where the construct appears to have little grammatical significance. They may be largely invisible to the analyst…There may be little that subjects do in such a language to attract the linguist’s attention, and as a limiting case, there may be nothing at all…I would say that the subject relation’s grammatical potential remains latent…Its invisibility does not entail its absence. (Langacker 1991: 320–321)

Clearly, Langacker would have to reject Diver’s austere but concrete starting-point in sound waves.

Rejection of the categories of sentence grammar is one major element of what differentiates Columbia School from other traditional and contemporary schools of linguistics. More than this, though, sets Diver’s theory apart from the others to which Columbia School is compared in the Introduction. Again, the differences stem ultimately from the views of the respective schools on the nature of theory. For the others, *theory* means starting out with some kind of *a priori* assumption or speculation. True, Jakobsonians and Guillaumeans have some different assumptions from traditional and generative grammar, and are thus justified in viewing their work as departures from traditional grammar. But Diver’s radically inductive, *a posteriori* view of theory sets Diver apart from them as well.

A case in point. The Introduction notes that Columbia School does not share the Jakobsonian interest in finding a universal model for all languages at all levels, phonology to grammar, which Jakobsonians attempt to achieve with binary features and markedness – Jakobson’s “six cardinal features”. This is a good point and an important point. However, it must also be made clear that the interest in universals is a very natural embodiment of a speculative, *a priori* approach to language, and that it is *this* which lies at the heart of the difference between the two schools. From the perspective of Diver ’95, how can one “set out to find” a universal model for all languages? One sets out to find whatever is there; and that may or may not turn out to be universal.

Note that it is this point, again, which turns the Columbia School analyst’s eyes first to morphemes and ultimately to the sound wave. There are really only these two ways to go when sitting down to analyze a language: one can start with the raw sound wave, or one can start with some sort of *a priori* classification. As Diver ’95 points out, if you come down anywhere in the middle, you have already made an arbitrary decision. The *a priori* route naturally leads to universals: if a set of linguistic categories exists prior to an individual analysis, then it is language-independent and pretty well has to be universal; and it is right here at this initial step that Columbia School parts ways with the others.
Again, the Introduction concentrates on details: Jakobsonian “markedness” and “search for semantic universals”. But we are offered no inkling of why one school should be interested in one thing and the other school not in that but in another thing. Such comparison, which does not get to the root of the disagreement, reinforces the impression some may have that what Columbia School linguists are doing is “soft” science, subjective, the same type of thing as literary criticism. There is a bigger picture behind these divergent interests that must be brought out for them to make sense.

The Introduction points out that Columbia School analyses employ such concepts as “exhaustively subdivided semantic systems”, “satellite relationships”, “oppositions of inclusion and exclusion”, a well-developed notion of the “interlock”, the “strategy”. But the impression is conveyed that these are a priori-type categories for Columbia School too, rather than constructs that achieved their place in the theory as the result of an insistence on fit between hypothesis and data.

Mention is also made of contrasting methods of validation. All sign-based schools, we are told, make use of short individual examples, sometimes decontextualized and sometimes invented. Only Columbia School relates meaning to macro-level discourse as well. This leads to the Columbia School use of counts. Once again, we are left wondering why this difference among the schools exists; its relation to a more general scheme, to a Main Idea, is not made clear. Readers may conclude that these differences are simply a matter of taste, like vanilla, chocolate and maple walnut.

Another important a priori assumption rejected by Columbia School is the generally-held view that linguistic meaning is compositional, the assumption that linguistically communicated messages can be segmented and each segment mapped to some element of linguistic form. The Introduction mentions this point as differentiating sign-based schools from generative grammar; but it is no less an important difference between Columbia School and the sign-based schools. The Columbia School view of meanings as sparse tools that significantly underdetermine messages (Diver 1975, 1995), which has been dubbed an “instrumental” view of meaning (Huffman 1997: 16 ff.), emerged as a result of analysis in the early development of the school and ultimately forced recognition of the pervasive influence of the human factor in the structure and functioning of language.

On this point we may compare Diver ’95 with an article by Hirtle (1995) in the same volume. Hirtle presents some of the fundamentals of the Guillaumean School through the vehicle of an analysis of the English morpheme -s, the traditional noun plural suffix. Citing the example: Books provide food for the mind, Hirtle says that the sense of the -s on books in this example can hardly be described as “more than one, plural”, but rather as “all, generic”, whereas he would accept the
meaning “more than one” in the example: *There were books all over the place.* With reference to this situation, Hirtle says:

Such polysemy is by no means rare for morphemes and words. On the contrary, a glance at a dictionary or a grammar suffices to show that it is in fact the general rule and as such constitutes an important part of the data for discerning the meaning potential of any such item. ... Being so widespread and so basic, the problem of polysemy calls for some general postulate concerning the nature of language meaning itself. Any approach which does not offer a solution to this problem at the outset tends to overlook, downplay, or even deny the existence of polysemy. ... the diverse senses of any item should be expressly sought out since they can each tell us something about the underlying meaning... accepting polysemy as one of the facts of linguistic life provides a starting point for analyzing the meaning of any item. (Hirtle 1995: 159)

Polysemy is here taken as an observational fact of language; the dictionary definition of a word, with its deliberate multiplication of “senses”, is taken as an accurate representation of meaning and, moreover, as part of the analytical problem. These are hallmarks of the assumption that linguistic meaning is compositional.

Hirtle postulates “discontinuate quantity” as one element of the meaning of -s. However, to account for the allegedly observable polysemy, it is necessary to posit also a formally stated “psychomechanical” process that allows a range of messages through greater and greater quantities to the final position corresponding to the maximum quantity, ‘all’. At the other extreme of this dynamic continuum the -s will cover even examples – like *a crossroads* and *a new airlines* – where a singular entity is referred to. For the -Ø of the so-called singular, Hirtle posits the meaning “continuity” along with the same psychomechanical process, a mirror image analysis of -s. This allows - Ø to be used for an entity which is in fact singular, as in *I took an aspirin*, as well as when there is a multitude on the scene, as in *Aspirin is an analgesic*. Looking at this in Columbia School terms, the process whereby one gets from meaning to individual messages is here being built into the linguistic encoding, the “meaning potential” itself; the mapping is thereby formally guaranteed.

Comparison of Hirtle’s analysis with Reid’s (1991) Columbia School analysis of the same forms is instructive. Reid does posit the meanings one and more than one for -Ø and -s respectively. Reid’s line of analysis would say that in *Books provide food for the mind*, we are in fact talking about more than one book, not just one; and in *Aspirin is an analgesic*, we are talking about a uni-form (sic) substance, whose potential division into multiple tablets is irrelevant for this message. By adopting an instrumental rather than a compositional view of meaning, Reid deals with Hirtle’s alleged counterexamples in a very different way: through the notions
of lexical choice, spanned opposition, and exploitation of meanings for messages of comment. Readers are referred to Reid (1991) for a full discussion.

It would be fascinating to continue the comparison of Columbia School to other linguistic schools from this perspective. Let us limit ourselves, though, to just a few brief and suggestive remarks comparing Diver '95 to generative grammar. First a quote from V. J. Cook's book *Chomsky's Universal Grammar*.

The goals of the theory are to describe language as a property of the human mind and to explain its source. To achieve these goals it establishes an apparatus of considerable complexity. Though the specific proposals put forward are not necessarily correct, the theory provides the unified framework within which they may be tested. (Cook 1988: 1)

This “theory” is surely coming from the pre-analytic, *a priori* end of the spectrum and contains, as part of the “apparatus”, many sub-theories: government/binding theory, x-bar theory, theta theory, case theory, and more. Each of these represents, not a solution to a problem, but a speculative framework for the application of a deductive model. In other words, Universal Grammar reaches for the big picture at the beginning and then tries to fit the details in with an analytical apparatus of ever-expanding complexity. Many contrasting quotes from Diver '95 come immediately to mind:

The term “theory” will here be used to refer to a summary of the general characteristics of successful solutions to individual problems. It definitely will not be used to refer to a speculative hypothesis that attempts to forecast in advance an as yet unattained solution to a problem. (Diver 1995: 45)

We do not want to take the position that, although we know very little as yet about language, the one thing we do know is that it has a very abstract and complex structure; for, obviously, until we are able to state what that structure is, there is no justification for saying there is any structure at all. (Diver 1995: 46)

Indeed, consider how many things are presupposed by Cook's statement: that one can have a theory without any correct specific proposals; that language is best characterized as a property of the human mind; that language has its source in some unspecified properties of the human mind; the notion “language” itself. For Diver, even this is something that might emerge at the end, not be assumed at the beginning:

...we probably want to save the English word “language” and apply it to the solution to the problem. We want to be able to say, at some point, “This, then, is what language is,” recognizing, in the word “language”, a common-sense reference to a certain kind of human activity, but not, again, a metaphysical reality whose properties we have to set out to discover. (Diver 1995: 46)
Diver himself (1975) pointed out two other initial assumptions of generative grammar that he rejected as instances of *a priori* reasoning: (1) the assumption that human language is unique to our species – that the study of the behavior of other species, and, in particular, of their communicative activities, and of the act of communication itself, will not illuminate the nature of human language; and (2) that language is also unique within the species, that it is unprofitable to attempt to understand human language through the study of other forms of human behavior. To this list can be added the assumption of a compositional view of meaning, and of course, that of the logic-based view of linguistic structure together with the categories of the Theory of the Sentence.

Otheguy (2002) has suggested that the important distinction between Columbia School and other theories is not merely a disagreement about whether these categories are or are not units of analysis; but rather, that generative grammar and other schools of sentence grammar take them as units of observation, whereas Columbia School considers them categories of an unsuccessful attempt at explanation. I would only add that Otheguy’s insightful formulation is but another facet of the Main Idea of Diver ‘95, the search for the starting-point of linguistics and its austere discovery in the sound wave, itself a facet of the radically inductive approach to theory building. Otheguy correctly points out that the very notion “categories of observation” makes no sense, that observations do not come in pre-packaged units, a point which corresponds to Diver’s recognition that there is only chaos at the bottom of it all.

4. Diver ’95 and explanation in science

Having examined Diver’s Main Idea as a contribution to the discipline of linguistics, I would like to suggest, a bit more diffidently, that the epistemology, the grounds for understanding in linguistics, represented in Diver ’95 may also offer a noteworthy contribution to science in general. This is another topic that would require a full-scale treatment, and, lacking sufficient expertise for such a treatment, I will here just point to a few ideas that I find suggestive of a wider applicability.

Diver espoused a unique conception of what constitutes *explanation*. He always began his courses at Columbia with a discussion of the structure of theory and his view of the meaning of explanation. Epistemological thinking is, even today, very much under the influence of the deductive model, in which laws – that is, assumptions or extrapolations from a limited set of observations – serve as premises, and one proceeds by axiomatic steps to derive specific observed phenomena. Explanation is thus equated with deductibility and, especially, with predictability. Such *a priori* reasoning is by no means disreputable among philosophers of science. The
deductive model provides people with a sense of security not tendered by Diver ‘95. In the words of Carl Hempel:

... the premisses of an inductive inference are often said to imply the conclusion only with more or less high probability, whereas the premisses of a deductive inference imply the conclusion with certainty. (Hempel 1966: 11)

But this notion of explanation, explanation in the sense of prediction, does not equal explanation in the sense of discovering causes of effects, Diver’s sense. Let me make this point clear with a parallel from another discipline: chemistry, in the development of the periodic table of the elements.

Nineteenth-century chemists had noted a periodicity in the properties of the elements. As the elements are ranked by atomic weight, certain sets of chemical and physical properties repeat themselves after every group of eight. Mendeleyev grouped all the known elements into a Periodic Table to represent these periodicities. There were some gaps in the table, and Mendeleyev predicted new elements would be found to fill these gaps. Those elements eventually were found and the predictions fulfilled.

The periodic table represents a heavy paradigmatic skewing. The “Periodic Law” of the elements is an extrapolation based on this table. Not only does it allow prediction of undiscovered elements. In several cases, atomic weights had been calculated which put elements in places that didn’t make sense in the table. New atomic weights were predicted on the basis of periodicity, the measurements were redone, and the predicted new atomic weights were found to be correct. Even more: several elements were placed out of order of their exact atomic weights in the table by Mendeleyev on the basis of their periodicity. When, later on, atomic numbers came to be used instead of atomic weights, the ordering of elements by atomic number fit Mendeleyev’s scheme, thus confirming the theory of atomic numbers.

As breathtaking as all this is, however, none of it explains the paradigmatic skewing, the periodicity. All of this work was done well before physicists began using spectral emissions, isotopes, radioactivity, and the magnetic properties of atoms to determine that periodicity is caused by the electronic structure (i.e. the arrangements of electrons) of individual atoms of the respective elements.

Thus, in the case of the Periodic Law, a law that was extrapolated from a limited set of preliminary observations achieved striking success as a predictor of atomic weights, atomic numbers, and even of undiscovered elements. However, the observational skewings captured by the law had an explanation that the law could not even begin to fathom. The law offered no actual understanding, no actual explanation.
Now, scientists of every kind, including linguists, would of course agree, in some way, with Diver’s insistence that hypotheses must fit observations. But Diver went further in developing the notion of an orientation and specifying the role orientations play in explanation. His development of this concept appears to be innovative and distinctive. The notion of an orientation describes a type of deductive control imposed on explanation by a motivating factor drawn from some area of knowledge, external to the discipline, which we believe we understand sufficiently for our immediate purposes. Diver’s orientations – the communicative factor, the human factor, the physiology of the vocal tract, etc. – prevent linguistic analysis from becoming mere deduction from premises.

Nor, on the other hand, is analysis an instance of what is sometimes called “narrow inductivism”, which, Hempel says:

... distinguishes four stages in an ideal scientific inquiry: (1) observation and recording of all facts, (2) analysis and classification of these facts, (3) inductive derivation of generalizations from them, and (4) further testing of the generalizations. The first two of these stages are specifically assumed not to make use of any guesses or hypotheses as to how the observed facts might be interconnected....

(Hempel 1966: 11)

Diver ’95, in contrast, says that hypothesizing begins immediately the moment we step away from raw observations. Furthermore, Diver specifically rejects the view that explanation is equivalent to generalization, what he calls “horizontal explanation”, the demonstration that a particular item is a member of a more general class of phenomena.

What is missing or murky in what Hempel describes is again the role of orientations. Such motivating factors are bypassed by reductionism, a form of generalization which is considered a valid type of deductive explanation; for instance, the notion that “physics is a more general science than biology”, “all of biology can be reduced to physics”, or, closer to home, “all of language can be reduced to neurobiology”. This kind of talk contrasts with Diver’s notion of a “discipline”: a body of explanations established by an appeal to a unique set of orientations. Mere reduction ultimately does not leave us any more enlightened about our observations than we were at the outset; in most cases, in fact, if it could actually be carried out (e.g. if someone could actually completely describe a linguistic event solely in terms of neural events, which is doubtful), it would result in such complexity that we would be considerably less enlightened. Diver’s idea was that it is profitable to set up explanation at various intermediate levels because of the availability of orientations in terms of which different aspects of linguistic phenomena can be understood. Thus, one set of orientations establishes the discipline we call grammar; a partially different set establishes the discipline we call phonology.
Unlike a law of classical science, an orientation, in Diver’s scheme, is not something to be explained by theory. We might contrast the two positions in the following way. In deductive explanation, “laws” serve as premises, and axioms generate phenomena. The explanandum – the body of observations which require explanation – comes out at the end; it is derived by a deductive process. In Diver’s view, the explanandum has to come first; the process has to start with it, otherwise there is no explanation.

5. Linguistic sign theory versus Columbia School

Finally, what are we to say of Diver ’95 as an instance of linguistic sign theory? According to the Introduction, the main idea is the sign; all sign-based schools accept the sign as the basic unit of *langue*. But Diver ’95 talks neither about the sign nor about *langue*. Its basic units are signals and meanings. This means that the sign as a unit doesn’t take on a life of its own, as it so often has in various lines of thought derived from Saussure; so that, for instance, a whole text may be called a sign. Rather, the pairing of signals and meanings merely serves as a control on positing minimal units of form and minimal units of meaning: you need each one to be sure you have the other.

The notion of “linguistic sign theory” is again theory in the *a priori* sense. We could say that these various schools share the notion of sign only if they all came up with the same signs; but of course they do not. Sign is therefore just another speculative notion; it is in no way a guarantee that all these schools are talking about the same thing. Perhaps, then, we need to recognize a new dichotomy in our taxonomy of schools of linguistics: Linguistic Sign Theory versus Diver’s *A Posteriori* Theory.

Columbia School linguistics had its origins at a time when generative grammar was first appropriating the spotlight for itself, and in a very flashy manner – accompanied by grand speculations about the human mind. Diver’s position was the polar opposite of generative grammar; it was not likely to, nor did it, attract many followers while the action seemed clearly to be at the other pole. It is natural for pioneering thinkers to reach for credibility by associating their ideas with those of recognized and distinguished predecessors. I believe that this partly explains why Diver emphasized his kinship with Saussure in the early days. But Diver was a maverick; he could not let somebody else do his thinking for him. Throughout his career, Diver set an example with his intellectual integrity and independence of spirit. Minority or no, he was not ashamed to insist that his ideas were distinct from those of others, and at the most fundamental level.
A great deal of what takes place in linguistics today is motivated by ideas and assumptions that conflict directly with the foundations of Columbia School theory. Much of this is enticing but ultimately sterile talk of theory-this and theory-that. Diver, too, had his use for the word *theory*, but what he intended was very different from what other users of the word intend. He wanted to make this difference clear, and this is why he highlighted that particular notion in “Theory”. Those of us who work in the Columbia School framework also have a duty to articulate those differences clearly and completely. Recent works such as those cited in this article, as well as other papers in the present volume, amply illustrate the directions where Columbia School is poised to make valuable contributions to the study of language. Now well advanced beyond its origins, Columbia School has defined a clear path for itself to follow. It can be guided more and more by the theory it has created.

Notes

1. This paper is based on a talk given at the Fifth Conference of the Columbia School of Linguistics, February, 1997. Thanks to Aaron Liebman for his many apropos suggestions on transforming an in-house talk into a published paper.
2. I owe this distinction to Joseph Davis (personal communication).
4. Note that there is no conflict here with Diver's notion of the deductive control imposed by orientations, discussed below. Orientations draw on conclusions worked out in other disciplines; we are discussing here the application of *a priori* decisions in the same discipline.
5. For example, Davis (1992, 2002).

References


CHAPTER 3

Phonology as human behavior

Inflectional systems in English

Yishai Tobin
Ben-Gurion University of the Negev

This paper summarizes the theory and methodology of Phonology as Human Behavior (PHB) (or Columbia School Phonology) and applies it to the inflectional morphology of English both synchronically and diachronically. The basic hypothesis is that inflectional morphology is both functional and frequent and should therefore be composed of phonemes that are unmarked or relatively easy to make. My second hypothesis is that this tendency for favoring unmarked phonemes in inflectional morphology should increase over time. I examine the phonological components of the inflectional morphology of Modern English and compare them with the phonological components of the inflectional morphology of Old and Middle English and then trace the parallel development of inflectional morphology in Proto-Indo-European to Proto-Germanic to Old English.

1. The theory: Phonology as human behavior (PHB)

This paper summarizes the basic theoretical and methodological tenets of the theory of Phonology as Human Behavior (PHB) (also known as Columbia School Phonology) (e.g. Diver 1979; Davis 1984 [1987]; Tobin 1997a); shows how it has been applied to various languages from diverse language families and to developmental and clinical phonology; and then applies it to the field of inflectional morphology in English both synchronically and diachronically. The theory of PHB, developed by William Diver and his students of the Columbia School, combines aspects of the “communication factor” inherent in Prague School phonology (Tobin 1988) with aspects of the “human factor” inherent in Martinet’s diachronic phonology (Martinet 1955). The major parameters of the theory are presented according to the functional semiotic definition of language as a sign system used by human beings to communicate. The fundamental axiom underlying the theory is
that language represents a struggle between the desire for maximum communication (the communication factor) with minimal effort (the human factor) (Tobin 1990a, 1994/1995). The major contribution of the theory of PHB is that it provides an explanation for the distribution within the speech signal: i.e. it tells us why the distribution of phonemes within a language is not random but motivated.

1.1 The four orientations underlying the theory of PHB

Four orientations underlie PHB:
1. the communication factor;
2. the physiology of the vocal tract;
3. the acoustic medium;
4. the human factor. (adapted from Diver 1979, Tobin 1997a)

The communication factor and the human factor have been introduced above and will be further elaborated upon in this paper. The physiology of the vocal tract is related to how human beings learn to control specific musculature to alter the air stream in the production of speech sounds (articulatory phonetics), and the acoustic medium refers to how human beings perceive these sounds (acoustic phonetics). These orientations are directly related to the semiotic and synergetic definition of language as a sign system used by human beings to communicate.

One of the aspects of the theory of PHB that separates it from other phonological theories is its recognition of the interdependence of the abstract phonological system of langue and the concrete distribution of sounds in parole. This integral connection between phonetics and phonology is supported by a theoretical and methodological model based on sound units similar to what have traditionally been called phonemes versus allophones (cf. Diver 1995: 62–72; Davis, this volume) that contain various features which systematically link them together according to their communicative function and reflect the characteristics of human perception, cognition and behavior relevant to their exploitation (Tobin 1997a: 18–23).

1.2 Viewing phonetics/phonology as human behavior

The following principles underlie the classification of sounds or phones into what are traditionally called phonemes and allophones. Phones should be classified as phonemes versus allophones based on: (1) their contribution to communication (phoneme versus allophone): the communication factor; (2) the specific primary distinctive versus secondary non-distinctive articulatory and acoustic features which speakers must learn: the human factor; (3) the relative non-predictability of the distribution of phonemes in minimal pairs (the communication factor) versus
the predictability of the complementary distribution of allophones (the human factor). There is a systematic interchangeability of phonemes and allophones across languages and within the same language in different periods that can be explained from the points of view of the communication and human factors. The fact that all languages have a similar number of phonemes (usually between 20–40) which are acquired in a similar order across languages and their diachronic versus synchronic relationships to alphabet systems and orthography can be related to both factors. The asymmetric relationship between the number of phonemes versus the number of allophones in language systems and our awareness of phonemes versus allophones can be explained from the synergetic point of view of maximum communication with minimal effort (Tobin 1997a: 18–23).

The following principles, derived from research on several languages, underlie the replacement of traditional and neo-traditional phonetic and phonemic categories and labels with new concepts related to PHB which illustrate the unity of sounds and sound classes as part of a single hierarchy (Davis 1984 [1987]; Tobin 1990b, c, 2000, 2002a-c; Azim 2002; Hameed 1999; Jabeen 1993; Fatihi 1987; Flores 1997; Dekker and de Jonge this volume; Tobin and Miyakoda 2001, this volume).

1. Consonants are referred to as phonemes of constriction.
2. Vowels are referred to as phonemes of aperture.
3. Features such as voiced versus voiceless, nasal versus oral, emphatic versus non-emphatic, ejective versus non-ejective, retroflexed versus non-retroflexed, palatalized versus non-palatalized, nasal versus oral, and labialized versus non-labialized, etc., are studied in terms of the number of sets of articulators being exploited simultaneously and the degree of the difficulty of their control from the point of view of the human factor.
4. Place of articulation for phonemes of constriction is replaced with the notions of active articulators versus passive receptors.
5. The tongue and the lips are viewed as active articulators for phonemes of aperture concerning the height and position of the former and the degree of rounding or spreading for the latter.
6. The division of the tongue into anterodorsum and posterodorsum as active articulators does not only serve to distinguish phonemes of constriction but can also replace the traditional categories of front versus back vowels respectively as the active articulators for phonemes of aperture.
7. Most (if not all) of the other various traditional categories of manner of articulation of consonants and vowels versus semi-vowels such as plosives, trills, flaps, glides, liquids, approximants, etc. should be replaced by a unified hierarchical set of degree of stricture, aperture and type of airflow that is applicable to
both phonemes of constriction and phonemes of aperture which now may be viewed as a single, holistic continuum or cline of sounds.

8. *Mobile* phonemes (e.g. stops which require the movement of the active articulators for closure and release or a trilled /ɾ/) versus *stable* phonemes (e.g. fricatives which hold the active articulators in a steady position or the lateral approximant /l/) can replace former categories of manner of articulation for phonemes of constriction and can be applied to both phonemes of constriction and phonemes of aperture replacing categorizations such as obstruents versus sonorants and syllabic versus non-syllabic phonemes.

9. Tenseness versus laxness and fortis versus lenis can be studied from the point of view of the amount of effort needed to control different sets of articulators and musculature for all sounds regardless of their classification as phonemes of constriction or aperture.

10. The difference between simple phonemes versus complex phonemes (stops and fricatives versus affricates and monophthongs versus diphthongs) can be explained from the point of view of the synergetic interaction between the desire for maximum communication with minimal effort.

The following principles can explain the non-random distribution of sounds both within the phonemic inventory of a language system as well as within meaningful units of language:

1. Maximum communication is achieved by investing human effort in the exploitation of the extremities of the oral vocal tract which provide the most distinct acoustic cues (e.g., grave and acute loci) in opposition to a neutralized center:
   a. the lips versus the posterodorsum in opposition to the apex (front-back-middle) the most adroit of the articulators for phonemes of constriction;
   b. the vowel triangle: /i/ versus /u/ in opposition to /a/ (front-back-middle) for phonemes of aperture.

2. When these optimal oral oppositions are exploited, further communicative distinctions are made by exploiting the musculature between the extremes and the middle or by extending their boundaries:
   a. by adding additional passive receptors (teeth, hard palate) or less adroit active articulators (anterodorsum, uvula, pharynx, glottis) or the vocal folds (for the glottal stop /ɬ/ and fricative /h/ as well as for tone, murmur, creakiness, etc. in addition to voicing) for phonemes of constriction and aperture;
   b. by exploiting the intermediary tongue positions between the high and low front and back points of the vowel triangle /e/, /ɛ/, /æ/, /ɔ/, /o/, etc.
3. Additional communicative distinctions can be made by exploiting the adroit active articulators in different ways requiring more effort and thus creating marked sounds such as lateral, retroflex, emphatic, ejective, velarized, nasalized, etc. phonemes.

4. More complex sounds requiring greater effort (affricates, diphthongs, etc.) can create further communicative distinctions as well.

5. In all of the above processes there also will be a favoring of adroit active articulators and a preference for the excitation of fewer sets of articulators when possible as a result of the human factor.

6. The differences in the higher communicative force of utterance-initial versus the lower communicative force of utterance-medial versus utterance-final positions also will affect the choice of more adroit versus less adroit, or more visual versus less visual articulators, and phonemes requiring one, two or three sets of articulators.

7. In most, if not all languages, the relative number and the proportion of marked versus unmarked phonemes will be similar, although the features being marked will differ from language to language.

8. In most, if not all languages, the number and kind of phonemes and their role in syllable structure will be based on the differences of the communicative forces of phonemes of constriction which impede the airflow in relation to phonemes of aperture which provide free movement of air.

1.3 The fundamental analytic position of PHB

The motivation for the explanation of the non-random phonological distribution is based on the following theoretical and methodological assumptions:

1. Users of a language behave as though they have learned certain distinctive units, the phonemes, which they deploy for communicative purposes.

2. We cannot observe directly what it is that they behave as though they have learned.

3. We can however observe the phonotactic skewing, a skewing that has been built up over the centuries and millennia in the very mouths of the speakers.

4. We can infer that these long-range skewings represent favorings and disfavorings on the part of users of the language. (It is to be observed that the skewings are not idiosyncratic to particular languages; their general characteristics recur from language to language.)

5. We can then examine the favorings and disfavorings against the background of the orientation -- which means with independent knowledge of what kinds
of favorings and disfavorings humans are prone to in areas other than the use of language.

6. We can infer that a disfavoring, for example, represents a difficulty in a learning process, and by a close examination of what it is that constitutes a difficulty in the way of a particular learning process, we can infer what it is that is being learned.

7. What it is that is being learned we may identify as a characteristic of the distinctive units. (adapted from Diver 1979; Tobin 1997a)

2. **Quantitative results and principles obtained from the theory**

All of the abovementioned research as well research on the phonotactics of certain classes of initial consonant clusters across languages (cf. Diver 1979; Davis 1984 [1987]; Tobin 2002b) yielded quantitative results, which support the following principles concerning the non-random distribution of phonemes in and across languages:

1. additional articulators are disfavored;
2. coarticulation by near articulators is disfavored;
3. coarticulation by the same articulators/phoneme is even more highly disfavored (particularly in the roots of Semitic languages);
4. different word (or root) positions have different communicative force and thus affect the favoring and disfavoring of different articulatory and acoustic features and phonemes;
5. apical articulations are favored in general and in final position in particular;
6. visual articulations are favored, particularly in word/root initial position;
7. explosive (mobile/stop) phonemes are favored in initial position;
8. turbulent (stable/fricative) phonemes are favored in final position;
9. transitions from one distinct constriction to another within a single phoneme (affricates) are disfavored;
10. consonant clusters are restricted concerning different articulatory and acoustic features (e.g., mobility / stability);
11. among constrictions, maximal constriction is favored;
12. among apertures, maximal aperture is favored;
13. sequences of phonemes with the same articulators are disfavored unless their juxtaposition is, by virtue of some other factor, mutually beneficial.

These principles empirically support the following conclusion:

**Conclusion #1**: Language in general – and phonology in particular – can be seen as a mini-max struggle: the desire to create maximum communication with minimal ef-
This synergetic principle is most evident in the nonrandom synchronic phonotactic distribution of phonemes in languages as well as in their diachronic development.

### 3. Developmental and clinical phonology: Natural phonology vs. PHB

In the last twenty years the theory of PHB has been applied to the areas of developmental and clinical phonology for monolingual and bilingual speakers of diverse languages (e.g., Enbe 2004; Enbe et al. 2006; Gan et al. 1995; Green 2005; Tobin 1995a, 1997a, b, 1999, 2002c; Tobin and Miyakoda 2001, this volume). This developmental and clinical research focused on both segmental phonology as well as on intonation and prosody highlighting the differences between PHB and other functional and generative theories of phonology such as Natural Phonology (e.g., Grunwell 1987; Ingram 1990; Stampe 1972 [1979]) and Optimality Theory (e.g., Kager 1999) which are compared and contrasted with PHB theoretically and methodologically in Tobin (2000). This research has shown that developmental and clinical phonology represent an even more extreme version of the mini-max principle underlying PHB found in Conclusion #1 above (cf. Tobin and Miyakoda, this volume).

### 4. PHB: From phonology to the lexicon

Tobin (2004) explored the connection between form and meaning in Hebrew roots. Specifically, he found a common semantic denominator in the tri-consonantal (CCC) Hebrew roots containing the phoneme /r/ as C-II (C-r-C) revolving around the general concept of *A Change in Structure* exemplified in the following examples of /p-r-C/ roots:

1. /p-r-m/ – ‘unstitch, undo stitches, untie, unbutton, rip; tear open, rend (clothes), rip, cut, chop’
2. /p-r-f/ – ‘fasten together (with a pin), pin up (hair), button’
3. /p-r-t/ – ‘expand, extend’
4. /p-r-t’/ – ‘divide into small parts; split, make change (‘break a dollar’); separate, specify, itemize, detail, divide’
5. /p-r-d/ – ‘separate, branch off, disintegrate, loosen, decompose, depart, divide, divorce’
6. /p-r-s/ – ‘slice (break, cut bread), spread, deploy, fan out, be divided, broken’
7. /p-r-S/ – ‘spread, cast, stretch out, expand, extend, unfurl’
8. /p-r-z/ – ‘exaggerate, overdo, overstate, be excessive, spread the fingers’
9. /p-r-ts/ – ‘break, break through, demolish; erupt, make a breach, crack, destroy; burst, rush upon; spread, increase, overflow, break out’
10. /p-r-r/ – ‘crumb, crumble, break into crumbs; shatter, undermine’
11. /p-r-q/ – ‘unload; deliver, set free, extricate; break, remove, dislodge, dismantle, disassemble, decompose, disjoint, demolish, liquidate, take apart, wind up, solve, resolve, crack’
12. /p-r-x/ – ‘crush, break, crumble; smash’
13. /p-r-h/ – ‘blossom, bloom, burst forth, burgeon, sprout; flourish, prosper, thrive, spread out’
14. /p-r-f/ – ‘uncover (esp. hair), be dishevelled, disordered, unkempt; wild-looking’
15. /p-r-f/ – ‘bring forth, bear, yield (fruit), bloom’
16. /p-r-h/ – ‘grow, produce, reproduce, be fruitful (fertile), thrive, flourish’

This general semantic field of *A Change in Structure* is reflected in this particular partial set of phonologically related roots in the following polaric processes:

a. *the division / separation of a whole into its component parts*: (1) /p-r-m/ ‘unstitch;... cut, chop’ (4) /p-r-t/ ‘divide, split, make change (‘money’)’; (5) /p-r-d/ ‘separate... disintegrate... decompose, depart, divide, divorce’, (10) /p-r-r/ ‘crumb, crumble; shatter...’; (11) /p-r-q/ ‘unload... break,... dismantle, disassemble, decompose, disjoint, demolish,... take apart, crack’, (12) /p-r-x/ ‘crush, break, crumble; smash’;

b. *the extension of the boundaries of a whole*: (3) /p-r-t/ ‘expand, extend’, (7) /p-r-S/ ‘spread, cast, stretch out, expand, extend, unfurl’, (8) /p-r-z/ ‘exaggerate, overdo, overstate, be excessive, spread the fingers’, (13) /p-r-h/ ‘blossom, bloom, burst forth, burgeon, sprout;... spread out...’; (15) /p-r-f/ ‘bring forth, bear, yield (fruit), bloom’, (16) /p-r-h/ ‘grow, produce, reproduce, be fruitful (fertile), thrive, flourish’; or:

c. *A Change in Structure* that includes both the antonymous or antithetical meanings of these polaric processes in a single root: (6) /p-r-s/ ‘slice (break, cut bread), spread, deploy, fan out, be divided, broken’, (9) /p-r-ts/ ‘break through, demolish; erupt, make a breach, crack, destroy; burst, rush upon; spread, increase, overflow, break out’; and/or

d. *a modification of the size, shape or proportions of a whole*: (2) /p-r-f/ ‘fasten together (with a pin), pin up (hair), button’, (14) /p-r-f/ ‘uncover (esp. hair), be dishevelled, disordered, unkempt; wild-looking’.

The analysis of all the /C-r-C/ roots found in various dictionaries showed that the general semantic field of *A Change in Structure* can be further divided into various related contextual sub-fields in the following descending order of the number
of /C-r-C/ roots they contain: Physical Changes (40), Extensions (general) (40), Extensions over Territory (20), Divisions (30), Border Changes (18), Heat Changes (12), Surface Changes (12), Mental State Changes (12), Connections (12), Control (10), Food/Health (8) Covering (7), Cloth (7) Collecting (3).

This connection between phonology and the lexicon was inspired by a general sign-oriented orientation (cf. Saussure 1916 [1959]; Jakobson 1941 [1969], 1970; Jakobson and Waugh 1979; Tobin 1988, 1989a; Waugh and Rudy 1991; and Andrews and Tobin 1996) and the principles of the Columbia School (Klein-Andreu 1983; Contini-Morava and Goldberg 1995; Contini-Morava and Tobin 2000; Reid, Otheguy, and Stern 2002) and PHB in particular (cf. Tobin 1989b, 1990a, 1991a,b, 1993, 1995b, 1996, 2002a-c) and is currently being expanded to include other triconsonantal root combinations as well.

5. PHB: From phonology to morphology

5.1 Inflectional systems in English

In this section, I will apply the principles of the theory of PHB to the realm of inflectional morphology in English both synchronically and diachronically. The basic hypothesis is that inflectional morphology (number, gender, case, tense, etc.) is both functional and frequent and should, therefore, in principle, be composed of phonemes which are unmarked or relatively easy to make. Our second hypothesis is that, diachronically speaking, this hypothesized tendency for favoring unmarked or relatively easier phonemes in inflectional morphology should increase over time. To verify these hypotheses I will first examine the phonological components of the inflectional morphology of Modern English and then compare them with the phonological components of the inflectional morphology of Old English and Middle English and then briefly trace the parallel phonological development of inflectional morphology in Proto-Indo-European to Proto-Germanic to Old English.

5.2 Modern English

Modern English is essentially an analytic language (Sapir 1921: 128), having a limited number of inflectional morphemes indicating grammatical concepts such as number, gender, case, tense, etc. The inflectional morphology of Modern English is listed below in Table 1:
Table 1. Inflectional morphology in Modern English

/Ø/ (zero) + /-s/-/z/ (sg./pl. nouns/verbs, genitive);
/-t/-/-d/-/i/ + /-in/ (past tense, past participle);
/-er/ + /-est/ (comparative, superlative)

The above inflectional morphology of Modern English may be described and explained according to the principles of PHB in the following way:

1. Suffixes are exclusively favored over prefixes.
   *Explanation:* Initial position (where the communication load is the highest) favors lexical morphemes while grammatical or functional inflectional morphology is reserved for final position (where the communication load is lowest).

2. The most favored suffix is composed of final consonants -C#.
   *Explanation:* Final consonants replaced final vowels as the most favored inflectional suffix in Modern English. Although consonants (phonemes of constriction) are harder to produce than vowels (phonemes of aperture) (the human factor) they provide clearer and more distinct communicative oppositions (the communication factor). This is also connected to an historical/phonological explanation: final vowels in unstressed syllables in English (a stressed-time language) are reduced to schwa /ə/ and thus lose their communicative distinction.

3. The less favored suffix is composed of -VC# syllables.
   *Explanation:* Fully stressed syllables (e.g. CVC) are not found in the inflectional morphology of Modern English but rather we find other unstressed –VC syllables when the final consonant of the lexical stem is the same (or a homorganic or a phoneme with a proximate active articulator) as the consonant found in the inflectional suffix (–C#). Thus coarticulation by near articulators or the same articulator in the form of a consonant cluster (human factor) which might also lead to a misunderstanding in communication (noun or verb ends with the same consonant as the plural or past inflectional suffix) (the communication factor) is doubly disfavored.

4. Final consonants made by the apex of the tongue are exclusively favored.
   *Explanation:* The apex of the tongue is the most flexible, sensitive and the easiest to control of all the active articulators and it has access to the most passive receptors.

5. Stops and fricatives (obstruents) are favored over nasals and other sonorants.
   *Explanation:* Obstruents allow for voiceless (one set of oral articulators) and voiced (two sets of articulators: oral + vocal folds) which allow allomorphs of the same inflectional paradigm to match the voicing or number of sets of articulators of the phoneme they are added to while sonorants are composed
of two (oral + vocal folds) and nasals of three (oral + vocal folds + uvulvar) sets of articulators which are more difficult to produce (the human factor) but provide more acoustic information and more distinct oppositions (the communication factor) but would change the number of sets of articulators for words ending in voiceless consonants. Here the human factor is stronger but accommodates the communication factor by the additional effort of adding an unstressed -VC# syllable.

6. There is an agreement in the number of sets of articulators being excited between the final consonant of the lexical stem and the final consonant in suffixes.

Explanation: It is easier to control the same number of the sets of articulators in proximate or adjacent phonetic environments than change the number of sets of articulators being excited in proximate or adjacent phonetic environments.

The so-called irregular inflectional morphology or morphological processes (such as internal vowel change or ablaut (sing/sang/sung)) today represent limited historical remnants of former grammatical inflectional systems which were probably semantically based and are now acquired lexically for frequently used lexical items rather than as grammatical systems (Tobin 1993: Ch. 12). The reduction of the various historical inflectional systems to favor the structure of ‘lexical stem + inflectional suffix’ has led to a limited number of suffixes composed of zero morphology or apical consonants or –VC# unstressed syllables containing apical consonants which match the voicing values of the consonant preceding them (the human factor): e.g. cat-s [kæts], dog-s [dɔgz] walk-s [wɔks], call-s [kɔlz], walk-ed [wɔkt], call-ed [kɔld], Pat-’s [pæts], Bill-’s [bɪlz]. This “linguistic economy” has resulted in the descriptive fact that some of the inflectional endings are homonymous: i.e. the same suffix appears in more than one system (e.g. /-s, -z, -1 z/ (sg./pl. nouns/verbs + genitive) and /-t, -d, -1 d / /-1d/, -m/ (past tense + past participle) which are perforce disambiguated by communicative function (part of speech) in the utterance. This concept of linguistic economy (the human factor) is broken when a consonant-only suffix has an allomorphic suffix composed of a syllable /-iz, -1d/. This allomorphic syllable suffix appears in the phonetic environment where the final consonant phoneme of the lexical stem is the same as the consonant phoneme of the suffix: bus-es [bʌsɨz], butt-ed [bʌtɪd], Jones-’s [dʒəʊnzɪz] or very similar phonetically and acoustically such as bush-es [bʌʃɨz] or witch-es [wɪtʃɨz]. The expansion of the consonant-only suffix to an unstressed -CV# syllable serves both to disambiguate the meaning of the stem with a singular noun ending with /s, z/ or a present tense verb ending with /t, d/ (the communication factor) and simultaneously avoids consonant clusters composed of the same phoneme, or homorganic phonemes, or phonemes using adjacent musculature (the human factor). The appearance of the same phoneme, or a homorganic phoneme, or a phoneme using an
adjacent musculature both in the final consonant of the stem and in the unstressed –VC# syllable can be justified by the mutual benefit found in the disambiguation of singular/plural past/non-past distinctions or the possible confusion created by similar sounding phonemes in final position.

To summarize: If one compares and contrasts the Modern English inflectional system with its historical counterparts one will find it to be the most economical or a finely honed example compared to Middle/Old English of the mini-max principle: (a) composed of suffixes favoring consonants (over unstressed vowels) and syllables (to disambiguate), (b) with the apex as the sole active articulator, and (c) a general agreement of voicing values that limits the number of gestures in the same phonetic environment to a minimum. The communication factor – in order to disambiguate potential breakdowns in communication – either adds suffixes, which are composed of syllables, and/or forces the addressee to contextualize homonymous suffixes according to their function or part of speech. In short, the Modern English inflectional system is a prime example of the axiom underlying PHB: the striving for maximum communication with minimal effort. The extent to which this mini-max principle has worked can be seen in the great distance between Modern English inflectional morphology and its Old English and Indo-European counterparts (cf. Borcier 1981: 84).

5.3 Old English

Old English is essentially a synthetic language (Sapir 1921: 128), having a large number of inflectional morphemes indicating grammatical concepts such as number, gender, case, tense, etc. The inflectional morphology of Old English (OE) differs from that of Modern English in the following ways:

1. OE has a tripartite number system composed of: singular, plural and dual.
2. OE has a tripartite gender system composed of masculine, feminine and neuter.
3. OE has a complex case system composed of nominative, dative, accusative, genitive, and instrumental.
4. OE declines the definite article and demonstratives according to number, gender, and case.
5. OE has weak and strong declensions for adjectives.
6. OE has a tense system composed of past/non-past.
7. OE has a mood system composed of indicative, subjunctive and imperative.

The inflectional morphology of Old English is listed in Table 2:
Table 2. Inflectional morphology in Old English

<table>
<thead>
<tr>
<th>Inflectional Morpheme</th>
<th>Declension/Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>/Ø/ (zero) /-(w)/-es/-e/ /-as/-a/-um/</td>
<td>nom-acc./gen./dat. + sg./pl. + masc./neu.</td>
</tr>
<tr>
<td>-ā- -jā- -wā-</td>
<td>dative plural–um</td>
</tr>
<tr>
<td>-u/-w/e/-e/- (w)a- or –(w)e/- - (w)a- or –(w)ena/- (w)um/-</td>
<td>nom/acc.-gen.-dat. + sg./pl. + fem.</td>
</tr>
<tr>
<td>-ō- -jō- -wō-</td>
<td>genitive plural–a</td>
</tr>
</tbody>
</table>

/Ø/ (zero) -a/-e/- /an/-an/-e/ /an/-an/-an/ /an/-an/-en/ /en/ /en/ /en/-um/-um/um/ | consonant declension: nom/acc./gen./dat. + sg./pl. + masc./fem./neu. + weak adjective declension |
| /Ø/ (zero) /u/ /-ne/-e/- /-es/-re/-es/- /um/-re/-um/ /-e/-e/- (a)/-a/- /-ra/-ra/-ra/-um/-um/-um/ | strong adjective declension: nom./acc./gen./dat/instr/ + sg./pl. + masc./fem./neu. |
| /-ra/-ost/-est/-ast/ | comparative/superlative |
| -e/-st/-d (or –t/ /að) | present indicative + 1/2/3 person + sg./pl. |
| /Ø/ (zero) /-e/-e/- /-est/- /-on/ | preterite indicative + 1/2/3 person + sg./pl. |
| /-e/-en/ | both subjunctives |

Even a rudimentary look at the list of inflectional morphemes in Table 2 indicates a large syncretism or overlapping of homonymous inflectional endings which can be seen in the noun-declensions:

1. Different forms have different functional loads: dative plural –um and genitive plural –a are common to all declensions.
2. Nominative / accusative plural (never distinguished in Old English) have a wide variety of forms including addition of syllables (stān/stānas [‘stone’], addition of different vowels (wite/witu [‘punishment’], hand/handa [‘hand’], vowel substitutions sunu/suna [‘son’], additions of various forms (lamb/lambru [‘lamb’], (guma/guman) [‘man’], substitutions (sunne/sunnan) [‘sun’], the use of an unmarked form (hūs, tungol, cynn, searu) [‘house’, ‘star’, ‘race/people’, ‘art/trick’], and ablaut, (bōc/bēc, mann/menn, frēond/friend [‘book’, ‘man’, ‘friend’].
3. There are eight functional slots for each noun declension; some have six distinct forms. For these eight potential communicative oppositions (wulf [‘wolf’]), others such as (guma) [‘man’] have only four distinct forms while sunu [‘son’]) only has three distinct forms, and different functions are distinguished in different words and word classes (cf. Bourcier 1981: 90).
All of the above indicate a less than economical and efficient distribution of inflec-
tional morphological forms.

The inflectional morphology of Old English may be described and explained
according to the principles of PHB in the following way:

1. Suffixes are exclusively favored.
   \textit{Explanation:} As above.

2. Suffixes are composed of final vowels and consonants -V#/C#, or VCV/VCC
   syllables.
   \textit{Explanation:} Both vowels (phonemes of aperture) and consonants (phonemes
   of stricture) are employed in the more complex inflectional morphology of
   OE, relative to ModE. Vowels are easier to produce (the human factor), but
   consonants provide clearer distinctions (the communication factor).

3. Suffixes composed of vowels are favored over suffixes composed of consonants
   or syllables.
   \textit{Explanation:} Vowels are easier to produce (the human factor) even though
   they produce less distinct communicative distinctions especially in unstressed
   syllables. In Mod E suffixes composed of relatively easy to produce apical
   obstruents with clearer communicative distinctions that can maintain the
   number of sets of articulators with the phonemes they are attached to now
   predominate over less distinctive but easier to produce vowels which are re-
   duced to schwa \textipa{/a/} in final unstressed syllables.

4. Suffixes with short vowels are exclusively favored over suffixes with long vowels.
   \textit{Explanation:} In Old English vowel length (short vs. long) was phonemic. Short
   vowels are unmarked and easier to produce (the human factor).

5. Final consonants made by the apex of the tongue and the lips are exclusively
   favored.
   \textit{Explanation:} Although consonants (phonemes of constriction) are harder to
   produce than vowels (phonemes of aperture) they provide clearer and more
   distinct communicative oppositions. The apex of the tongue followed by the
   lower lip are the most flexible, sensitive and the easiest to control of all the
   active articulators (the human factor) and the lip has an additional visual ad-
   vantage (Tobin 1997a).

6. Nasals and other sonorants (including vowels) are favored over stops and fric-
   tives.
   \textit{Explanation:} Although obstruents allow for voiceless (one set of oral articula-
   tors) and voiced (two sets of articulators: oral + vocal folds) while sonorants
   are composed of two (oral + vocal folds) and nasals of three (oral + vocal folds
   + uvulvar) sets of articulators that are more difficult to produce (the human
   factor) they provide more acoustic information and more distinct oppositions
(the communication factor). In Mod E suffixes composed of a small set of relatively easy to make apical obstruents have generally replaced a large variety of vowels and nasals and other sonorants and combinations thereof of various degrees of difficulty to produce which had different functional loads were distributed in a rather inefficient and seemingly random way based on a large degree of syncretism.

In Old English the so-called irregular inflectional morphology of ablaut was more frequently used and probably more systematic (i.e. might be considered to be part of the grammar and acquired as such rather than as lexical exceptions). Although Old English has a more complex inflectional morphology system than Modern English in the amount of forms, in their number of functions, as well as in their phonological composition as stated above, there is still a great deal of syncretism, overlapping and homonymy in the system (cf. Table 2) which makes it less than ideal from the point of view of the communicative factor as can be seen in the OE case system summarized succinctly and exhaustively by Bourcier (1981: 143):

Even more crucial for the future of the language was the loss of case-distinction. Loss of the accusative was hardly surprising: already in Old English this case had been formally indistinguishable from the nominative not only with neuters and with all plurals but also with many singulars (including the large ā-stem group with all its sub-types, as well as the ī-stems and the the ū-stems), so that noun inflexion has ceased to be a principal means for distinguishing between subject- and object rôles. The datives, singular and plural, had on average enjoyed far clearer formal distinction than had accusatives; yet by Chaucer's time that whole case too had virtually disappeared. This disappearance, fully comprehensible only in light of the accusative/dative syncretism affecting the pronoun …, had reflected general loss of case-feeling, partly due to the availability of prepositions for expressing syntactic relationships, together with the variability of post-prepositional case-usage; in an Old English phrase such as on þissum ēare the inflexions were not only ‘redundant’ but expendable, because the same sense can be clearly and more economically conveyed by in this year. Both losses, originally determined by syntactic conditions, entailed further adjustments on that level. (Bourcier 1981: 143)

In contrast with Modern English, the Old English system of inflectional morphology does not provide an optimal compromise between the struggle for maximum communication with minimal effort and therefore has changed rather drastically. Future changes in Modern English are always possible but will probably be less drastic.
5.4 Middle English

Middle English represents a transition period between Old English, an essentially complex *synthetic* language, and Modern English, an essentially simple or economic *analytic* language. The basic changes in the inflectional morphology from Old English to Middle English are the following:

1. The OE tripartite number system (singular, plural and dual) was reduced to a bipartite system (singular and plural).
2. The OE tripartite gender system (masculine, feminine and neuter) was lost morphologically with a subsequent loss of feeling for gender.
3. The OE complex case system (nominative, dative, accusative, genitive, and instrumental) was syncretized to nominative + accusative/dative (oblique) with a subsequent loss of feeling for case.
4. The OE declensions for the definite article and demonstratives were lost.
5. The OE weak and strong declensions for adjectives were lost.
6. The OE tense system (past/non-past) and mood system (indicative, subjunctive and imperative) developed into systems of auxiliaries and periphrastic verb tense, aspect and mood systems.

The inflectional morphology of Middle English is listed in Table 3:

<table>
<thead>
<tr>
<th>Sym</th>
<th>Morphology in Middle English</th>
</tr>
</thead>
<tbody>
<tr>
<td>/Ø/ (zero) /-e/-es/ /-es/ or (-y/-is)</td>
<td>all noun declensions + nom.-acc./gen. + sg./pl. + all genders</td>
</tr>
<tr>
<td>/Ø/ (zero) /(-e)</td>
<td>all pronoun/article/adjective declensions + nom./acc.-dat./gen. + sg./pl + all genders</td>
</tr>
<tr>
<td>/Ø/ (zero) /-ed/ /-en/</td>
<td>present indicative + 1/2/3 person + sg./pl.</td>
</tr>
<tr>
<td>/Ø/ (zero) -(e)de-</td>
<td>preterite indicative</td>
</tr>
</tbody>
</table>

The above inflectional morphology of Middle English may be described and explained according to the principles of PHB in the following way:

1. Suffixes are exclusively favored. 
   *Explanation:* As above.
2. Suffixes are composed of final vowels –V# and –VC# or -(V)CV# syllables. 
   *Explanation:* A similar but less complex syllable structure as in OE above.
3. The consonants in inflectional suffixes are made by the apex of the tongue exclusively.

   *Explanation*: The apex of the tongue is the most flexible, sensitive and the easi-
est to control of all the active articulators and it has access to the most passive
receptors.

4. The consonants in the inflectional suffixes are fricatives and stops (ob-
struents).

   *Explanation*: Obstruents allow for voiceless (one set of oral articulators) and
voiced (two sets of articulators: oral + vocal folds) and stable phonemes (frica-
tives) are favored in final position.

Middle English represents the period of “ablaut-leveling” whereby the more fre-
quently used and probably more systematic grammatical *ablaut* system of Old Eng-
lish became more lexical or “irregular” as in Modern English. Middle English also
represents a wholesale or across-the-board leveling of inflectional morphology for
number, gender, and case for all parts of speech and serves as a transition period
between synthetic Old English with a complex inflectional morphology and ana-
lytic Modern English with a simpler and more streamlined or economical inflec-
tional morphology. There are several ways of explaining this leveling of inflectional
morphology summarized succinctly and exhaustively by Bourcier (1981: 142):

   Why the Old English declensional system had suffered such drastic reduction
   can be understood from examining more closely and noting its weaknesses. Sev-
eral paradigms had depended mainly on distinctive vowel-timbers threatened by
weak stress. None had catered for certain basic distinctions, such as that between
nominative and accusative plural. Subsequent developments illustrate how all the
levels of language – sounds, inflexions and word-order – constantly interact with
one another. Once unstressed vowels had all been reduced to [ə], distinctions
based on distinctive timbers ceased to exist, so that their surviving nouns had to
be reformed according to other patterns; similar *declensional transfers* had been
taking place already in Old English, with the -ō- declension taking over many
feminine -i- stem nouns and the -ā- declension many masculines and neuters
from both the -i- and -u- stems. Simultaneously, shifts in *case-feeling* (clearer
with pronouns than with nouns …) were reducing the number of inflexional dis-
tinctions required. Increasing uses of existing alternative methods of marking
syntactic rôles, that is of set word-order and of prepositional phrasing, allowed
these inflexional losses to proceed almost unhindered. Then in their turn the in-
creasing inflexional losses accelerated the shift towards analytic syntax…. So, by
the Early Middle English period a few dominant inflexional patterns, themselves
subject to simplification, were ousting those no longer viable either because of
sound-change or of rarity (sound-changes apart, languages burdened with multi-
ple sets of equivalent inflexions often evolve towards more economical systems).
Survival of one pattern rather than another must have depended partly on auditory distinctiveness and partly on numerical predominance, and on both counts the stān-declension with its clear and resistant -s endings, was fitted for the long-term success it was to enjoy at the expense of all its rivals. (Bourcier 1981: 142)

It is clear that the relatively unmarked phonemic composition of the inflectional morphemes adhere over time from Old English to Middle English to Modern English and have been further streamlined by the motivation for maximum communication with minimal effort as outlined above according to the principles of PHB.

5.5 From Proto-Indo-European to Proto-Germanic to Old English

It is evident from the historical development of the inflectional morphology of English that the phonological component (or the form/signifiant) of the inflectional linguistic signs developed along the pattern that was hypothesized: i.e. became more economical (was reduced to a significant favoring of apical obstruent consonants which matched the voicing values of the final consonants preceding them in the lexical stem). In those cases where syllables appear rather than consonant-only suffixes, the addition of a syllable was motivated by a potential confusion or breakdown in communication. Inflectional morphology composed of (short) vowels only (in unstressed syllables and therefore reduced to schwa /ə/ as English developed into a stress-timed language) were lost. Similar findings supporting the hypotheses presented in this paper predicting that (a) the phonological component of inflectional suffixes would be composed of unmarked phonemes or of phonemes that are relatively easy to produce, and (b) this tendency would increase over time, are also found when one examines nominal declensional systems from Proto-Indo-European (PIE) to Proto-Germanic (PG) to Old English (OE) (Bourcier 1981: 87):

1. In those cases where PIE and PG have a syllabic suffix consisting of -VCV#, OE has a –CV# suffix.
2. In those cases where PIE and PG have syllabic inflectional suffixes consisting of -(DIPH)-V-C# or -(DIPH)/VC#, OE has a suffix composed of a final –V# only.
3. In those cases where PIE and PG have a vowel-only suffix, the –V# is a phonemic long vowel while in OE the cognate vowel-only suffix is a phonemic short vowel.
4. PIE, PG and OE have both apical and labial consonants in inflectional suffixes while Middle and Modern English have only apical consonants.

The selected inflectional morphology for PIE, PG and OE is listed in Table 4:
Chapter 3. Phonology as human behavior: Inflectional systems in English

Table 4. Inflectional Morphology in Proto-Indo-European → Proto-Germanic → Old English

<table>
<thead>
<tr>
<th>PIE</th>
<th>PG</th>
<th>OE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ā</td>
<td>-ō</td>
<td>-u</td>
</tr>
<tr>
<td>PIE</td>
<td>-ā-m</td>
<td>PG</td>
</tr>
<tr>
<td>-ō-m</td>
<td>-ō-s</td>
<td>OE</td>
</tr>
<tr>
<td>-e</td>
<td>-e</td>
<td></td>
</tr>
</tbody>
</table>

- [ā- declension: nom./acc./gen. + sg. + fem.]

<table>
<thead>
<tr>
<th>PIE</th>
<th>PG</th>
<th>OE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ou-es</td>
<td>-au-iz</td>
<td>OE</td>
</tr>
<tr>
<td></td>
<td>-au-i</td>
<td>OE</td>
</tr>
<tr>
<td></td>
<td>-a</td>
<td></td>
</tr>
</tbody>
</table>

- [u- declension: gen./dat. + sg.+ masc.]

<table>
<thead>
<tr>
<th>PIE</th>
<th>PG</th>
<th>OE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-on-s</td>
<td>-an-s</td>
<td>OE</td>
</tr>
<tr>
<td></td>
<td>-on-m</td>
<td>PG</td>
</tr>
<tr>
<td></td>
<td>-an-um</td>
<td>OE</td>
</tr>
<tr>
<td></td>
<td>-an</td>
<td>OE</td>
</tr>
<tr>
<td></td>
<td>-an</td>
<td>OE</td>
</tr>
</tbody>
</table>

- [weak consonant declension: nom./acc./gen./dat. + sg.]

This paper presents a first step in the examination of inflectional morphology in a limited number of languages. Similar analyses based on PHB for the derivational morphology of Hebrew (Shokty and Perelshtein 2005) has recently shown similar initial findings for derivational morphology as well. Based on this preliminary study one can make the following tentative conclusions:

**Conclusion #2**: The non-random phonotactic distribution of inflectional morphology may be viewed as another example of the mini-max struggle: inflectional morphology represents functional linguistic forms which are used frequently and therefore favor (unmarked) phonemes which require less control over the articulatory tract or mechanisms.

**Conclusion #3**: The non-random distribution of inflectional morphology may be viewed as an example of the linguistic economy or the synergetic relationship found in the linguistic sign: the simpler the meaning of the linguistic sign (signifié) the more unmarked the phonemes / morphemes of its signal (signifiant) the more complex the meaning of the linguistic sign (signifié) the more marked the phonemes/morphemes of its signal (signifiant) (adapted from Tobin 1989, 1990a, 1991a,b, 1993, 1995b, 1996, 2004).

These conclusions may have wider theoretical and methodological implications regarding the classification of language systems from the point of view of the synergetic or semiotic economy of language systems in general and in their acquisition and retention in particular as stated in the following conclusion:

**Conclusion #4**: The non-random distribution of inflectional morphology may be viewed as an example of a synergetic, linguistic, or semiological economy: in general, inventories of items are more readily learned, digested and recalled when their items are sub-categorized in a preliminarily intuitive way in the form of a mnemonic economy or a memory-saving device – i.e. a way of cutting down the
amount of effort required for memorizing and remembering all the well-formed elements of a system... If... the objects to which the figurae [signals] bear some similarity are closely connected with the semantic content of the signa [signs] in whose expression these figurae are used, the interrelatedness achieved by this similarity will also make the memorization of the signa (and their appropriate form-meaning associations) considerably more economical. (Adapted from Hervey 1982; 1985: 141, 155; 1988: 27, 32).

6. Summary and conclusions

In this paper the theory of PHB was extended to the realm of inflectional morphology synchronically and diachronically to see whether the principles of the theory can explain the favoring of unmarked or relatively easy to make phonemes in inflectional morphological signs and their non-random distribution within these signs. This research in English can be related to analyses of other languages representing diverse language families.

In a diachronic study of Old Church Slavonic, Old Russian and Modern Russian, Buk (2003) found that there was a phonological and morphological simplification in general and in inflectional suffixes in particular in each subsequent stage of the languages: (1) the number of phonemes was reduced, (2) suffixes became shorter, (3) the short-long phonemic distinction in vowels disappeared, (4) the oral-nasal phonemic distinction in vowels was lost, (5) /æ/ changed to /ɛ/ and the merging of /ɛ/ either with /e/ or /i/, (6) the loss of markers for hard and soft variants of consonants (/b/, /b/), (7) complex palatal units /ʃtʃ/ and /ʒdʒ/ and affricates /dʒ/ and /dzʒ/ were depalatalized or simplified, (8) reduction of /a, o, e/ into schwa /ə/ in unstressed syllables, (9) increase in preference for apical consonants.

From the point of view of PHB the following principles can be applied to inflectional suffixes which appear in word-final position where the burden of communication is lowest: (1) the preference for vowels over consonants, (2) the preference for front vowels over back vowels, (3) the preference for hard consonants over soft (palatalized) consonants, (4) the favoring of mobile stops with maximum stricture over stable fricatives with lower stricture, (5) the disfavoring of additional articulators, (6) apical phonemes are favored (for Modern Russian), (7) transitions from one distinct constriction to another in a single phoneme (deaffrication) (for Modern Russian). The morphological inflection systems also became simpler over time: (1) fewer verb conjugations and noun cases, (2) a tripartite number system (singular, dual, plural) was reduced to a bipartite system (singular, plural), and (3) the tense system was simplified.
Similar results indicating phonological and morphological simplicity in general and increasing over time were also found for the following languages from different language families: Latin (Cohen 2001) compared with modern Romance languages such as Ladino, Spanish, and Portuguese (Tubul 2002; Oron 2003), Arabic (Saif 2004), Hebrew (Oren 1997; Shokty and Perelshtein 2004) and Hungarian (Salmon 2002). Additional research for derivational morphology is being planned and executed in these languages and for both inflectional and derivational phonology for other languages to see whether there is a similar non-random distribution of phonemes synchronically and diachronically in inflectional and derivational affixes which further support the principles of the theory of PHB. Preliminary results in Hebrew (Shokty and Perelshtein 2005) show similar tendencies for derivational morphology as well.

Notes

1. The reader may note that the suffix -ing is not included in the above list of inflectional suffixes. I collected the inflectional forms from various histories of the English language and was surprised myself that none of them included -ing. I reached the conclusion that -ing may be viewed as being inflectional (a participle as opposed to other verb/adjectival participle forms in a paradigm) but it is also used as a gerund with a nominal use – i.e. changes part of speech allowing it be used as a subject as opposed to a verb/adjectival participle function which could also classify it as “derivational”. I therefore followed the lead of the various history of English books I used, concentrating on the more “classical” paradigmatic inflectional categories such as number, case and historical tense (versus aspectual forms) and adjectival comparative/superlative forms. Even if -ing were included in my analysis, it would not really alter the results in a significant way other than instead of the “exclusive use” of the apex or obstructs as opposed to nasals as final consonant suffixes – it would be the “overwhelming majority” except for [ŋ] which is posterodorsal (but often pronounced as apical [n]).

References


CHAPTER 4

Phonological processes of Japanese based on the theory of phonology as human behavior

Yishai Tobin
Ben-Gurion University of the Negev

Haruko Miyakoda
Tokyo University of Agriculture and Technology

By analyzing speech errors (normal and pathological) and loanwords of Japanese within the theory of Phonology as Human Behavior, we seek to account for why processes such as substitution occur as they do by referring to the “struggle” between speakers’ desire for maximum communication (the communication factor) and minimal effort (the human factor). We conclude that (1) the error patterns or the processes observed in loanword adaptations are not random but motivated and that clinical phonology represents a more extreme version of the “mini-max” struggle where the human factor overrides the communication factor; (2) the communicative forces found within different word positions have a great influence on how and where the phonological processes of loanwords occur.

1. Introduction

In this paper, we will analyze the phonological processes of Japanese within the theory of Phonology as Human Behavior (PHB). In particular, we will focus on the processes observed in speech errors and loanwords. By employing data from Japanese, we will demonstrate that the theory of PHB can account for phonological processes in this language: why the sounds or structures are altered as they are in speech errors and loanwords.
2. The theory of phonology as human behavior

In this section, we will limit ourselves to the concepts of PHB that have direct relevance to our discussion in the following sections. For details, refer to Diver (1979), Tobin (1997a), and Tobin (this volume).

2.1 General concepts of PHB related to phonological processes

The semiotic act of communication can be seen as a “mini-max” struggle: the desire to create maximum communication with minimal effort. The encoder and the decoder are the opposing two positions involved in this mini-max struggle. The general tendency is for the encoder to attempt to communicate based on minimal effort; however, minimal effort on the part of the encoder will place a heavier burden on the decoder’s inferential abilities. From the standpoint of the decoder, therefore, in order to achieve maximum communication, maximal effort on the part of the encoder is desirable because there is less chance of a breakdown in communication. This playoff between encoders and decoders becomes prominent especially in speech communication (this point has been emphasized in other theoretical frameworks such as Optimality Theory, cf. Prince and Smolensky 2004: 3). To illustrate, consonant cluster reduction (e.g. floor > for, step > tep), an often observed process in different stages of language acquisition of English is, according to PHB, due to the human factor: a consonant cluster requires greater effort to produce than a consonant-vowel sequence, and may be reduced or replaced at the expense of maximum communication. To take another example, assimilation, another often observed phonological process in various languages, is assumed to take place due to the human factor: assimilation is influenced by, or dominated by, a sound that is produced with fewer articulatory gestures (e.g. a non-velar/nasal/labial sound changes to a velar/nasal/labial because velar/nasal/labial sounds involve fewer articulatory gestures: duck > guck, top > bop).

Although the principles related to the human factor can account for many of the phonological processes observed, it is also important to take into consideration the communication factor, which puts emphasis on the desire for maximum communication. For example, according to the statistics presented in Paradis and LaCharite (1997), when changes are made to loanwords, adaptation of segments is largely preferred over deletion. Since deletion is an “easier” process than adaptation (substitution) in terms of speech production, we would assume that deletions would be preferred in such cases. But out of the 16,000 tokens, 78.5% involved adaptations while only 2.5% involved changes through deletion of a segment. This fact shows that speakers not only attempt to minimize their effort in speech production, but at the same time try to maximally preserve the original word struc-
ture. This is due to the communication factor. Similar examples can be attested in developmental phonology. For example, it has been pointed out in the literature that substitution occurs more often than deletion where communicative force is stronger. In the case of English, consonants in stressed onset positions rarely undergo deletion but instead substitution takes place (e.g. tiger > kiger, but not ‘iger). This is because the initial stressed position of words has the highest load of communication (it pays to work harder, so to speak, in that position). This is unlike the case for word-final position. Here, deletion, not substitution, is preferred because the position at the end of a word has less communicative force compared to the initial position (e.g. out > [au], bike > [bai]).

A phonological theory must be able not only to describe, but also to explain this compromise. The theory of PHB is one such theory that encompasses the struggle between both the encoder and the decoder. In combining the communication factor with the human factor, PHB claims that a more complete theory of phonology should and must be able to take both the communication and the human factor into account together. In terms of Optimality Theory (OT), this can be stated as the conflicting forces observed between the constraints of markedness and the constraints of faithfulness. Markedness for OT represents the pressure for the prevalence of unmarked types of structure (in PHB terms, the more natural, less complex, less difficult human factors). Faithfulness represents the preservation of lexical contrastiveness that allows more distinct and clear communicative distinctions and oppositions, i.e., in PHB terms, the communicative factor (Tobin 2000 compares and contrasts PHB and OT on these and other issues as well).

2.2 Principles of PHB

Although the principles of PHB were initially employed to explain the nonrandom distribution of phonemes in various languages, it has further proved to be effective in accounting for the nonrandom distribution of developmental and clinical data (see Tobin 1997a, 1997b, this volume).

In the next section, based on speech error data from Japanese, we will demonstrate that developmental and clinical phonology both represent a more extreme version of the “mini-max” struggle where the human factor overrides the communication factor.

3. Aphasia and PHB

Some researchers and clinicians have assumed that the errors in aphasic speech both within individuals and across the different classifications of aphasia are basi-
cally random and inconsistent (Tobin 1997a: 290). One of the main reasons for this lies in the fact that many differences have been found in the speech of patients suffering from different kinds of aphasia as well as in that of individuals diagnosed as having the same kind of aphasia.

As we shall demonstrate below, many of the generalizations made about the speech of aphasics can be described and explained from the viewpoint of PHB. In addition to the principles listed above, the following additional principles (1, 2) were established by analyzing aphasic data according to the theory (Ani 1995; Bar-Lev 1995; Tobin 1997a: Ch. 10).

(1) Principles involving the human factor:
   a. More errors are made in consonants than in vowels.
   b. Aphasics appear to have more difficulty producing consonants in word-initial position (initial position: harder to articulate).
   c. More errors are made in marked structures (requiring greater control over the musculature of different sets of articulators simultaneously) than in unmarked structures (e.g. errors are observed more in consonant clusters compared to consonant-vowel sequences, hence the ideal CVC syllable).
   d. The more frequently a phoneme is used in the language, the lower the relative frequency of errors for that phoneme (favorings represent ease of articulation) (e.g., /a/, the most frequently used vowel (Stemberger 1992: 518), is rarely involved in errors in Japanese).

(2) Principles involving the communication factor:
   a. There are usually more errors of substitution (although articulatorily deletion is easier, substitution takes place in order to preserve communicative oppositions).
   b. Aphasics have more difficulty producing word-initial position than word-final position.
   c. There is a tendency to replace phonemes with other phonemes that differ in one distinctive feature only rather than in more than one (e.g. back consonants substituted by apical consonants usually preserve the same manner and voicing values: k > t book [bʊt]).

Regarding (2b), previous research in clinical phonology based on PHB as well as other theories has shown a general favoring of errors (particularly substitution errors) in word initial position, especially among stutterers, aphasics, and the hearing-impaired (Tobin 1995, 1997a, 1997b). The interaction between the communication and human factors is also reflected in the differences of the distribution of phonemes in different word positions. Although phonemes of constriction (i.e.,
consonants) are more difficult to produce than phonemes of aperture (i.e. vowels), the communicative distinctions they create are clearer. Therefore, there are more consonants than vowels in all languages, and the most frequent syllable structure generally observed crosslinguistically is CV(C). It is also one of the first structures that appear in language acquisition, as indicated in the learning path of syllable types given in Figure 1.

\[
\begin{align*}
A: & (5)CVCC, VCC > (6)CCV, CCVC \\
& (1)CV > (2) CVC > (3)V > (4)VC > (7)CCVCC \\
B: & (5)CCV, CCVC > (6)CVCC, VCC
\end{align*}
\]

(Levelt and van de Vijver 2004: 209)

Figure 1. Paths of acquisition of syllable types

Although two variant sub-orders (A, B) were observed, all the Dutch learning children initially, i.e. at stage (1), acquired the CV type, and the validity of establishing this type as the starting point of the learning path has been supported by data from other languages as well. In word-initial position, where the burden of communication is the highest, we have an almost random distribution of phonemes of constriction still reflecting their different relative degrees of difficulty with a possible slight favoring of visual phonemes (i.e., labials and some apicals) which are perceived both aurally and visually. Therefore, initial position with a high communicative force also has the strongest possibility to appear with consonants of all degrees of difficulty, making it a prime target of articulatory errors.

Developmental and clinical phonology can be viewed as an extreme version of the synergetic mini-max struggle: the human factor very often overrides the communication factor. This is due to either extreme minimal effort or a lack of control over the articulatory tract or mechanisms. For example, in the case of substitution processes, speakers are often observed to employ the readily available phoneme already found in their repertoire. However, at the same time, the factors that facilitate communicative oppositions may also play an important role. Examples of such principles of PHB that were established based on developmental and clinical data can be summarized as in (3):

(3) a. The preservation of as many distinctive features as possible in substitution processes, which require more effort than deletion processes.

b. The preservation of as many communicative oppositions as possible in the original word, for example, the number of phonemes per word, in substitution processes.
c. The preservation of the original phonetic structure of the word in deletion processes not involving syllable reduction and in reduplication (e.g. reduplication often is triggered as a means to avoid more difficult sound combinations and/or to maintain the number of syllables in the word: *noodle* [nunu]).

d. The preservation of the stressed syllable bearing the most communicative information if the original structure of the word is reduced by the deletion of syllables (e.g. *banana* > [nænæ]).

3.1 Japanese speech errors based on PHB

The material forming the basis of this study is a collection of 501 spontaneous speech errors in spoken Japanese collected by the second author (Miyakoda 2002a, 2002b, 2003; Tobin and Miyakoda 2001).

The speech error data from normal speech were compared with 106 pathological errors of five aphasic patients (CY, HO, KT, HS, SM). The majority of the pathological errors were collected by trained speech pathologists in therapy sessions with free conversation and naming tasks (Miyakoda 2003).

The errors were analyzed within the framework of PHB. The error patterns are accounted for by employing the principles maintained within the theory presented above.

3.2 Results and discussion

In the following, (4)–(14), we will list our findings accompanied by an explanation derived from the principles of PHB.

(4) In both normal and pathological errors, there are more errors involving consonants than vowels (principle 1a).

Japanese provides a good testing ground for this claim, because consonants and vowels appear in the spoken language in fairly even proportion (consonants: vowels = 0.81: 1).

In our Japanese data, both normal and pathological errors showed a strong preference to involve consonants. In the case of normal speech errors, 46.9% involved consonants, compared to just 19.4% for the vowels. In the pathological data, 50% of the errors involved consonants, and only 7.5% involved vowels. The result is summarized in Table 1.
Table 1. Errors: consonants vs. vowels

<table>
<thead>
<tr>
<th></th>
<th>Normal speech errors</th>
<th>Pathological speech errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>involving consonants</td>
<td>235 (46.9%)</td>
<td>53 (50%)</td>
</tr>
<tr>
<td>involving vowels</td>
<td>97 (19.4%)</td>
<td>8 (7.5%)</td>
</tr>
</tbody>
</table>

What is particularly striking about this result is that, although vowels occur slightly more than consonants, more errors are observed for consonants than for vowels, showing an “anti-frequency” tendency. We may interpret this tendency as being related to the difference between consonants and vowels in controlling the musculature of the vocal tract. That is, consonants (phonemes of constriction) are more difficult to produce than vowels (phonemes of aperture) because they require greater control over the musculature of the vocal tract. Because they are harder to produce, they are involved in errors more than vowels. But consonants also produce more clear-cut distinctions than vowels. They are worth the extra effort to enhance communications.

(5) In both the normal and pathological data, substitution errors occurred the most.

If we consider the fact that the three major processes in errors are substitutions, deletions, and additions, we may evaluate each process from the point of view of the communication and the human factors. Deletions represent both the easiest process with the least communicative contribution. Additions (epentheses) require more effort but usually ease transitions between articulators and may not contribute to, or even hinder communication. Substitutions, however, require effort but maintain the structure of words by preserving the number of phonemes in words, thus contributing to better communication. In Tobin (1995, 1997a), a tendency for substitutions to appear in phonetic environments bearing high communicative force (word initial position and stressed syllables) and for deletions to appear in phonetic environments with low communicative force (word final position and unstressed syllables) was found in both functional, developmental and pathological, clinical data across languages and individual speakers.

Among the 235 consonant errors in normal Japanese speech, 220 were of the substitution type. Within the 97 vowel errors, we observed 91 substitutions. As for the pathological data, all errors of both the consonantal and vocalic type were substitutions. The result is depicted in Table 2.
Table 2. Substitution errors

<table>
<thead>
<tr>
<th></th>
<th>Normal speech errors</th>
<th>Pathological speech errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>involving consonants</td>
<td>220 (43.9%)</td>
<td>53 (50%)</td>
</tr>
<tr>
<td>involving vowels</td>
<td>91 (18.2%)</td>
<td>8 (7.5%)</td>
</tr>
</tbody>
</table>

Although there was no tendency for the normal speech errors to favor word-initial position, more errors were observed in word-initial position in the pathological data (cf. principles 1b, 2b above). Within the 106 errors, 61 were of the substitution type, and among these errors, 31 (50.8%) involved word-initial position.

(6) In normal speech errors, the vowel /a/ often becomes the target in vowel substitution errors (principle 1d).

As mentioned above, out of the 97 vowel errors, there were 91 substitutions. Details are given in Table 3. Rows indicate the phonemes found in the original form, and the columns show the phonemes that appear in the error form.

Table 3. 91 Substitution vowels

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>i</th>
<th>u</th>
<th>e</th>
<th>o</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>i</td>
<td>1</td>
<td></td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>u</td>
<td>1</td>
<td>6</td>
<td></td>
<td>1</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>e</td>
<td>16</td>
<td>5</td>
<td>2</td>
<td></td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>o</td>
<td>13</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>12</td>
<td>15</td>
<td>13</td>
<td>20</td>
<td>91</td>
</tr>
</tbody>
</table>

The ranking for the “target-prone” vowels is as follows: /a/ (31 errors) > /o/ (20 errors) > /u/ (15 errors) > /e/ (13 errors) > /i/ (12 errors). This result is in accord with the claim presented by Jakobson (1941 [1972]) regarding the issue of substitutability: there is a tendency for the unmarked category to take the place of its marked counterpart. In terms of PHB, we may interpret the unmarkedness of /a/ by reference to its ease in production: the low central neutral vowel /a/ is the easiest vowel to produce, which provides us with maximum aperture. However, no tendency could be observed in the pathological data.

(7) According to Jaeger (1992), in the case of English child phonology, the feature most violated is [place], and the least violated is [nasal] as shown in the following hierarchy (principle 3a): Place > Continuant > Fricative > Voice > Nasal.
In our data of normal children, the feature violated the most is [place] (active articulators), and the least violated features are [nasal], [voice] (number of sets of articulators): Place > Continuant > Nasal > Voice.

In the case of pathological speech errors, the feature most violated is [place], and the least violated is [voice] (numbers in parentheses show the number of errors for each feature): Place (27) > Nasal (13) > Continuant (12) > Voice (6). This hierarchy clearly resembles the one for the Japanese child phonology given above.

The reason why place features (active articulators) are involved in errors frequently is that children and aphasic patients have to learn to control the muscles to activate the various active articulators. In general, front consonants are easier to articulate than back consonants, hence, the assumption that back consonants such as /k/ are marked and universally liable to be replaced by front consonants such as /t/ pervades the literature (Beckman et al. 2003: 21). This implies that the easiest to control are the front articulators such as lips and the apex, and it becomes harder the further back the articulators are.

(8) In consonant substitution errors, the feature most violated is [coronal] (apical).

Among the 220 consonant substitution errors (normal speech), we found that 29 errors involved apicals being replaced by other places of articulation, followed by 20 velars and 18 labials replaced by other places of articulation. In the case of pathological errors, out of the 27 errors involving place, all 27 (100%) involved apical consonants. Apical consonants are produced with the apex of the tongue (the most flexible and sensitive articulator). They are the easiest to use of all the active articulators and the most frequently used in all languages. Therefore, because of their frequency, they may also become involved in errors the most.

Regarding apicalization, the normal speech data show that errors from apical to apical are observed the most: Apical > Apical (92 errors), Non-apical > Non-apical (54 errors), Apical > Non-apical (29 errors), Non-apical > Apical (27 errors), Others (18 errors).

Since apicals are considered to be the most unmarked consonant class, we would predict that the substitution errors would target towards these segments, just as in the case of vowels as seen above (where the most unmarked /a/ was the most “target-prone” among the five vowels). However, contrary to our prediction, the apicals do not replace other places of articulation, but rather, they are replaced by others. This tendency is observed not only in speech errors but in assimilation processes as well. Let us take as our example Sino-Japanese place assimilation. In assimilation processes of Sino-Japanese, word-final apicals assimilate in place to following word-initial consonants, but in the case of nonapicals, assimilation does
not occur, but instead the vowel /u/ is epenthesized to avoid a heterorganic consonant cluster. Examples are given below in (9) (Cho 1990: 55–56):

(9) Sino Japanese place assimilation
   a. Apicals assimilate in place of articulation
   
   bet + taku > bettaku ‘detached villa’
   bet+ kaku > bekkaku ‘different style’
   bet + puu > beppuu ‘separate cover’

   b. Nonapicals epenthesize /u/
   
   gak + see > gakusee (*gassee) ‘student’
   gak + cyoo > gakucyoo (*gacyyoo) ‘school president’
   gak + mon > gakumon (*gapmon) ‘learning’

Word-final apical /t/ assimilates in place to following word-initial consonants as in (9a), whereas word-final noncoronal /k/ does not assimilate to following word-initial consonants (9b), but instead the vowel /u/ is epenthesized, in order to avoid a heterorganic consonant cluster.

A similar tendency has been reported in Stemberger and Stoel-Gammon (1991) based on speech errors of English.

(10) The elements involved in a phonemic speech error have similar phonemic environment.

C > C
   karini > [t]arini ‘if’ (apicalization)
   tozansha > to[d]ansha ‘mountain climber’ (z>d, maximum stricture)

V > V
   ondoku > ond[a]ku ‘reading out loud’ (o>a, maximum aperture)
   CV > CV
   kekkon > k[a]kkon ‘marriage’ (e>a, maximum aperture)

In phonological processes, the original structure of the word is maintained as much as possible (preservation of as many distinctive features, communicative oppositions, syllables, etc. as possible) (principles 3a, 3b). The same goes for English errors as indicated in (11):

(11) C > C

   left hemisphere > heft lemisphere
   rolling pin > rolling pill

V > V

   the boy bows > the bay bows
   Bev and Bill > Biv and Bell (Laubstein 1987: 343)
However, there are exceptions to this correlation in Japanese, and these involve the moraic phonemes depicted in (12):

(12)  

\[
\begin{align*}
\text{V} & > \text{C} \\
\text{kei zai teki} & > \text{ke[N]} \text{ za[N]} \text{ teki} & \text{‘economical’} \\
\text{CV} & > \text{Q} \\
\text{o ku sa n} & > \text{o [Q] sa n} & \text{‘madame’} \\
\text{C} & > \text{V} \\
\text{ran kin gu} & > \text{ra[i] kin gu} & \text{‘ranking’}
\end{align*}
\]

Japanese is a language with a very restricted phonological inventory. The syllable structures allowed in Japanese include CV (/ta/ ‘field’), V (/ii/ ‘stomach’), and CCV (/sjo/ ‘calligraphy’). In practice, the mora will often overlap with the syllable, and indeed in many cases moras are syllables. The three syllable types just referred to are all one-mora syllables. However, Japanese has several elements that serve as an independent unit (mora). Although they do not qualify as an independent syllable, these elements are positioned so as to follow the basic syllable structures CV, V, and CCV. The resulting form is one syllable consisting of two morae. These elements, the so-called moraic phonemes, fall into four kinds: the nasal coda (N), the geminate consonant (Q), the second half of long vowels, and the second half of diphthongal vowel sequences.

The moraic phonemes were initially introduced to the Japanese language through borrowings from Chinese. Examples of words with moraic phonemes are given below in (13):

(13)  

\[
\begin{align*}
a. & \quad \text{nasal coda (N): ro n do n ‘London’} \\
b. & \quad \text{geminate consonant (Q): pi Q tsu ba a gu ‘Pittsburg’} \\
c. & \quad \text{second half of long vowels: ro o ma ‘Rome’} \\
d. & \quad \text{second half of diphthongal vowel sequences: ta i ‘Thailand’}
\end{align*}
\]

The nasal coda represents the sounds [n], [m], [ŋ], depending on the following phonetic context (e.g. pa N to [panto] ‘bread and; paN mo [pammo] ‘bread also; paN ga [panga] ‘the bread’). The geminate consonant represents a doubling of the consonant that appears in the onset of the following syllable (e.g. kiQ sa [kissa] ‘tea room; ki Q te [kitte] ‘postage stamp; kiQ pu [kippu] ‘ticket’). These moraic phonemes, although not capable of forming an independent syllable, do serve as an independent timing unit just as the regular CV one-mora syllables. The syllable structure of Japanese can be classified into two groups: the one-mora syllable types (CV, V, CCV) and the two-mora syllable types (CV, V, CCV followed by N, Q, or the second half of long vowels or diphthongs).
Figure 2 shows the internal structure of the syllable of Japanese.

Although the mora is a subsyllabic unit, it is this unit and not the syllable which is important in dealing with various phonological processes of Japanese. In the case of moraic phonemes, consonants can be replaced by vowels, and vice versa. This is allowed just as long as they interact within their own kind.

The behavior of these moraic phonemes can be accounted for by the syllable position hypothesis. This hypothesis, initially put forth by Boomer and Laver (1968), entails that in speech errors, phonemes in initial syllabic position replace those in initial position, nuclear replace nuclear, and final replace final, thus maintaining the position within the syllable. Simply put, this hypothesis accounts for why consonants interact with consonants only, vowels with vowels. Although this claim has been supported in many studies (cf. MacKay 1970; Laubstein 1987), lately some researchers have started to question the validity of this hypothesis (cf. Meijer 1997). These researchers suggest that rather than “syllabic” similarity, the reason why consonants and vowels do not interact with one another in speech errors is due to their “phonetic” similarity.

The examples of the moraic phonemes strongly support the syllable position hypothesis. If, as some researchers claim, it were the “phonetic” similarity that is crucial, we would not expect an interaction of consonants being replaced by vowels (and vice versa) to occur. What distinguishes the moraic phonemes from the other phonemes is their position within the syllable: moraic phonemes are the only ones that can occupy coda position. Similarly, evidence from pathological speech and child phonology given in (14) support this claim (Miyakoda 2003):

(14) a. Moraic errors in aphasic speech

\[
\begin{align*}
C > V & \quad V > C \\
\text{zjo[se]} & > \text{zjo[seN]} \text{ ‘female’} & \text{sha[se]} & > \text{sha[seN]} \text{ ‘to sketch’} \\
\text{[naN]do} & > \text{[nai]do} \text{ ‘closet’} & \text{[teN]do} & > \text{[teN]do} \text{ ‘degree’}
\end{align*}
\]
b. Moraic errors in child phonology
C (moraic) > C (moraic)
[daN]go > [dak]ko ‘dumpling’
[peN]gin > [pep]pin ‘penguin’
C > V
[paN]da > [pa:]da ‘panda’
[waNwaN] > [wo:wo:]’doggie’
V > C
[ke:]ki > [kik]ki ‘cake’
(hi)[ko:]ki > [kok]ki ‘airplane’

In terms of PHB, the special behavior of the moraic phonemes may be accounted for by the communication factor (preservation of original structure: principle 2a). That is, as long as the prosodic structure of the original word is maintained, the coda position of the syllable may be filled in by any one of the four types of moraic phonemes, whether it be consonants or vowels.

In the next section, we turn our attention to loanwords in Japanese.

4. Japanese loanwords within PHB

When a language incorporates a foreign word into its system, the typical sound process that takes place is to substitute the sounds that do not exist in that language with those that do belong. In the following section, we will be dealing with Japanese loanwords with focus on the communication factor. We will demonstrate that not only the human factor (ease of pronunciation), but also the communication factor plays an important role in determining how loanwords are incorporated into the Japanese language system.

4.1 Sound alternation

Substitution is the typical process that takes place in loanwords. This sort of sound alternation is not only observed in Japanese but is widely acknowledged as the most basic process in many of the languages of the world (Tobin and Miyakoda 2004). This can be accounted for in terms of PHB as due to the principle related to the human factor, that is, in substitution processes, the speakers use the readily available phoneme already found in their repertoire. Examples are given below in (15):

(15) a. /ʃ/ > [Φ]
  fork > [Φo:ku]
  knife > [naiΦu]
b. \(/\nu/ > [b]\)
   - *vanilla* \([\text{vәnилә}] > [\text{banира}]\)
   - *vacation* \([\text{вәkeyʃәn}] > [\text{bakeʃоN}]\)

c. \(/\theta/ > [s]\)
   - *three* \([\thetari:] > [\text{suri:}]\)
   - *theory* \([\text{θиәри}] > [\text{seori:}]\)

d. \(/\delta/ > [z]\)
   - *breathe* \([\text{bri:ð}] > [\text{buri:zu}]\)
   - *that* \([\text{ðæt}] > [\text{zaQto}]\)

Another point worth mentioning is that in substituting, there is preservation of as many distinctive features as possible (for example, cf. \(/\theta/\) vs. \(/s/\), \(/z/\) vs. \(/\delta/\) which differ only in one feature, \([\text{distributed}]\), \(/b/\) vs. \(/\nu/\) differing only in \([\text{continuant}]\): principle 2c).

### 4.2 Resolving Consonant Clusters

In the case of Japanese, another typical process that takes place in incorporating loanwords is to insert a vowel in between the CC clusters. This is generally assumed to take place due to the syllable structure of Japanese: the basic syllable unit is CV, hence a vowel is inserted to resolve CC sequences. However, in terms of PHB, consonant cluster sequences require greater effort to pronounce and are considered to be “more difficult” compared to usual consonant-vowel sequences. Therefore, even in languages that allow CC clusters, the CC sequences are reduced or replaced at the expense of maximum communication.

If a language has consonant clusters, they often appear in initial position where the force of communication is the highest. In word-final position, where the force of communication is the lowest, there is a strong statistical preference for phonemes or clusters of phonemes requiring the least amount of effort to produce: e.g. vowels, or voiceless, apical phonemes of constriction (Tobin 2002: 197). Some examples of vowel insertion in consonant clusters are given below in (16):

(16) *Christmas > kurisumasu*
    *stress > suitoresu*
    *thrill > suriru*

Vowel insertion can also take place at word final position. By inserting a vowel, the closed syllables are transformed into open syllables as in (17):

(17) *pass > pasu*
    *salad > sarada*
This again can be accounted for by assuming that the vowel is inserted in order to change the syllable structure that adheres to the Japanese language. However, in the case of moraic nasals, no insertion takes place, as shown in (18):

(18) nasal phoneme
tempo > teNpo
can > kaN

In terms of PHB, this is due to the principle in (3b), which states that the original phonetic structure of the word is preserved as much as possible (the communication factor). Even though for most Japanese, the pronunciation would be facilitated if the structure of the words were altered to conform to the basic syllable structure CV, the force to preserve the original structure becomes active at the expense of the human factor.

In the following sections, we will take a look at some of the more extreme cases where the communication factor plays an important role in accounting for why deletion occurs in some loanwords. As we have previously stated, deletions represent the principle of least effort (the human factor) often at the expense of the communication factor. In the case of loanwords that are adapted into the native vocabulary, the process of deletion (the human factor) trumps the communication factor because the original foreign word may not even be known by native speakers, nor is it crucial in their understanding and use of these new words in their lexicon and speech. Anecdotal evidence of this can be found in the speech of Israelis who add the Hebrew masculine plural suffix -im to English loanwords with the -s plural suffix: e.g. brakes-im for ‘brakes of a car’ or Eskimos-im for ‘Eskimos’. Other cases of English loan adaptation can be found in examples such as bildi-taNur ‘oven’, where bildi is an adaptation of the English term built-in.

4.3 Deletion processes in loanwords

Originally, most Japanese words were bimoraic. It was only after the influence of the Chinese language that longer words came to enter the language system. Still, the average word-length is three to four morae. Due to this limitation in length, long foreign words are made into shorter forms by deleting some element from the original word.

4.3.1 Deletion of final consonants

In (19) we have examples of loanwords that have deleted the final consonant of the original word. In terms of PHB, this is due to the fact that communicative force is weakest at word-final position. Therefore, even if the final consonant is deleted, this would not affect communication.
Deletion of word-final syllables

In (20) we have examples of deletion of word-final syllables (morae). Here again, deletion takes place in the word-final position due to its weak communicative force.

\[(20)\]

amateur > ama
building > biru
location > roke

Deletion of suffixes and articles

All the words listed in (21) are examples where the functional elements of the original words or phrases have undergone deletion. This may be due to the degree of communicative information that function words have: since functional words themselves do not bear much communicative information, the communicative contrast would not be affected much even if they underwent deletion. Therefore, the human factor trumps the communication factor, and the functional elements are dropped to allow easier speech production.

\[(21)\]

a. \{s\}
   cornflakes > [ko:NΦureiku]
pajamas > [paʤama]
slippers > [surıQpa]
stockings > [sutoQktNgu]
sunglasses > [saNgurasu]
three strikes > [suri:sutoraiku]
b. \{'s\}
   women’s lib > [u:maNribu]
   fielder’s choice > [Φi:ruda: tfoisu]
lamb’s wool > [ramu u:ru]
c. \{ed\}
   condensed milk > [koNdeNsu miruku]
corned beef > [ko:N bi:Φu]
dried fruits > [doraιΦuru:tsu]
reserved seat > [riza:bu jito]
old fashioned > [o:rudo ΦaQ∫oN]
d. \{ing\}
   *measuring cup* > [meʤa:kaQpu]
   *happy ending* > [haQpi: endo]

e. \{ment\}
   *engagement ring* > [eNge:dʒi r1Ngu]

f. \{the\}
   *off the record* > [oΦureko]
   *on the air* > [oN ea]

4.3.4 *Deletion of word-initial position*

So far, we have found only one example where the word-initial position of the original word is deleted: *American* > [meriken]. This word differs from the previous examples in two ways: first, the word-initial element, not the final, is deleted, and second, the reason for this deletion process taking place is due to the stress assignment of the original word, and not due to the communicative force of word position. Here, the word-initial vowel is deleted because it does not bear stress. This is in accordance with the principle which states that the stressed syllable is preserved due to its communicative force.

4.3.5 *Deletion in compound words*

When compounds are incorporated into Japanese, the initial mora of the two elements (usually two morae) is usually retained and the rest deleted, as in (22). Compared to word-final position, the initial position has more communicative force, hence, is retained.

(22) *sexual harassment* > [sekuhara]
    *word processor* > [wa:puro]
    *plastic models* > [puramo]

5. Conclusion

In this paper, we have analyzed speech errors and loanwords of Japanese within the theory of PHB. We have found that both the human and communication factors account for why the processes observed occur as they do. In particular, we have learned that in Japanese, the different communicative forces found within different word positions have a great influence on how and where the phonological processes occur in this language.
Although very little work has been done on the concept of syllables and morae in PHB, we have shown in this paper that the theory can account for the behavior of moraic phonemes in Japanese.

Acknowledgements

We would like to thank Fukuko Shusse for providing us with the aphasic data. The work reported on in this paper was supported by a grant from the Japan Society for the Promotion of Science (No: 14710326).

References

This paper applies the theory of Phonology as Human Behavior to an analysis of the distribution of consonants in monosyllabic words in Byelorussian. Like other languages, Byelorussian shows a direct connection between the effort that speakers make to control the active articulators, involved in the production of phonemes, and the favorings or the disfavorings of these phonemes in various phonotactic distributions.

1. Introduction

This paper presents an analysis of the non-random phonotactic distribution of consonants in monosyllabic words in Byelorussian according to the theory of Phonology as Human Behavior (PHB). PHB is grounded in a sign-oriented view of language as a system of systems used by human beings to communicate. Communication is based on the production of distinct sounds – the phonemes – through the control of the musculature of the vocal tract. The “human factor” implies the relevance of human intelligence, memory limitations, and human efficiency that reflect the nature of human beings who constantly look for maximum communication with minimal effort. See Tobin (1995: 7, 1997, this volume) for further discussion. The theory was first developed by William Diver (1979) in an analysis of initial consonant clusters in English. Since then, its findings have been extended to the non-random combinations of vowels and consonants in English and in language in general (Diver 1993, 1995) as well as to other languages such as Italian (Davis 1984 [1987]), Hebrew (Tobin 1997, 2002), Urdu (Azim 2002; Jabeen 1993), Mewati (Fatihi 1987), and Spanish (Flores 1997). The theory has also been applied to developmental and clinical phonology, dialectology, and to a wide range of historical, psycholinguistic and sociolinguistic issues (Tobin 1997, 1999, 2000, 2002).
The goal of this paper is to apply the principles of PHB to Byelorussian. The data for this study consist of all the monosyllabic words appearing in Арфаграфичны слоўник [Арфаграфічны слоўник] ‘The Byelorussian Spelling Dictionary’ (Loban and Sudnik 1978) and in Руско-беларускі словарь [Русско-белорусский словарь] ‘The Russian-Byelorussian Dictionary’ (Grabtshikov 1990). The use of a dictionary as a database was established by Davis (1984 [1987]: 6) as the optimal way for counting each lexical entry regardless of any syntactic, pragmatic, historical or stylistic constraints. Moreover, a lexical approach “permits an evaluation of the forces that operate phonologically within the word, largely independent of semantic content.” The choice of monosyllabic words avoids concomitant factors that could influence the distribution of phonemes such as syllable stress and vowel reduction in non-stressed syllables.

2. Traditional phonemic inventory of Byelorussian

Padluzny and Cekman (1973) provide one of the best known analyses of the sounds of Byelorussian. They investigate the Byelorussian phonetic and phonological system in comparison with the systems of other Slavic languages by using a variety of tests, mainly X-ray photography and palatograms. The great importance of their study is in their techniques and methods of classification, used to establish an inventory of Byelorussian phonemes. Table 1.1 presents an example of the traditional phonemic inventory of Byelorussian (for other traditional analyses of Byelorussian, cf. Avanesov 1978; Biryla 1978; Cekman 1969; Martynau and Padluzny 1975; Miller 1982; Podluznyj 1969; Sudnik 1972; Sussex 1974; Vyhonnaia 1986).

Table 1 reveals a relatively complex system of consonants with a less complex system of vowels. Byelorussian is rich in obstruents (stops, fricatives and affricates) which like nasals and the lateral /l/ have palatalized counterparts. Most obstruents appear also in voiced-voiceless pairs. At the same time, a very small number of consonants do not have voiced or palatalized counterparts.

The correlation of voice [in Contemporary Byelorussian] [I.D.] is distinctive for most segments (only j, r, l and the nasals are excluded); the correlation of palatalization plays a more modest functional role since the post-dental fricatives and affricates lack palatalized counterparts. (Wexler 1977: 180)
Table 1. Traditional Phonemic Inventory of Byelorussian

<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th>Dental</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>p</td>
<td>b</td>
<td>p'</td>
<td>t</td>
<td>d</td>
</tr>
<tr>
<td>Nasal</td>
<td></td>
<td>m</td>
<td>m'</td>
<td>n</td>
<td>n'</td>
</tr>
<tr>
<td>Fricative</td>
<td>f</td>
<td>v</td>
<td>f'</td>
<td>s</td>
<td>z\̊</td>
</tr>
<tr>
<td></td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>\̃</td>
<td>h\̃</td>
</tr>
<tr>
<td>Affricate</td>
<td>ts</td>
<td>ts'</td>
<td>dż\̊</td>
<td>t\̃</td>
<td>dż\̃</td>
</tr>
<tr>
<td>Lateral</td>
<td></td>
<td></td>
<td></td>
<td>l\̄</td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td></td>
<td></td>
<td></td>
<td>r</td>
<td></td>
</tr>
<tr>
<td>Semivowels</td>
<td>w**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semivowels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Close Front</td>
<td>i</td>
<td></td>
<td>(y)\̋</td>
<td>(u)</td>
<td>u</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(o)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td>a</td>
</tr>
<tr>
<td>Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Here and below by superscript \(^{1}\) I will understand palatalized phonemes according to traditional classification.

** The phonemes /u/ and /i/ are used in the initial and postconsonantal positions. When preceded by a vowel, they change into diphthongs, whose second component is /w/ or /j/, respectively. Compare, for example, yłjo [u`ju] 'I weave' versus cařjo [saw`ju] 'I will have woven', ičći [is`tsi] 'to go' versus bońcji [`vyjstsi] 'to go out'.

\(^{1}\) There are different opinions on the allophone 'ý/ /y/ (presented also as /i/ in traditional phonetic transcription) in Slavic languages. According to Fradkin (1986 [1985]:viii), /y/ is "a back, unrounded allophone of i". My opinion is that /y/ is rather an anterodorsal allophone of /i/ that is used after non-palatalized consonants.
Wexler provides a thorough diachronic study of the phonological system of Byelorussian from its emergence from Common Slavic to today. He presents the Byelorussian phonological system but does not explain why specific groups of consonants are “defective”, i.e. do not have voiceless or palatalized counterparts. Moreover, Table 1 shows that palatalized counterparts are relevant for most consonants, which means that the functional role of palatalization in Byelorussian is far from being “more modest”, as Wexler says. By contrast, appealing to the human and communication factors provides a possible explanation for the “defectiveness” of these phonemes. We have learned from the phonological studies of other languages (cf. Diver 1979; Davis 1984 [1987]; Tobin 1997, 2002, etc.) that the reuse of the same musculature is disfavored across languages, and the same holds for Byelorussian as well. Among non-palatalized consonants in Byelorussian, there are fricatives /ʃ/, /ž/ and affricates /tʃ/ and /dʒ/. They are produced by excitation of the active articulator (the blade of the tongue, or the anterodorsum) in the direction of the passive receptor (the soft palate). To palatalize these already palatal articulations, we would have to make a disfavored articulation: to reuse the same musculature, the anterodorsum.

We have also learned that combinations of similar gestures (likes) are favored and combinations of different gestures (differents) are disfavored within and across languages. The same holds for Byelorussian as well. This principle accounts for the trilled /ɾ/ that is also non-palatalized in Byelorussian. /ɾ/ is articulated by the active vibration of the apex of the tongue raised to the alveolar ridge. To palatalize /ɾ/, we would have to both raise the anterodorsum retroflexively in the direction of the soft palate and keep it relatively stationary during the vibration of the apex. This combination would represent a simultaneous stable-mobile gesture in a close proximity and would require a greater effort to be produced. Therefore, it does not exist in Byelorussian. It is also disfavored in other Slavic languages. In Russian, for example, /ɾ/ is palatalized. Nevertheless, even a cursory glance at Russian dictionaries shows that in Russian there are many more entries beginning with non-palatalized /ɾ/ (e.g. /ɾa/, /ɾo/, /ɾu/, /ɾy/) than with its palatalized counterparts (/ɾjə/, /ɾe/, /ɾjə/, /ɾo/, /ɾu/).

The reasons for the apparent exception of the non-palatalized stops /t/ and /d/ can be found in the history of Byelorussian. The palatalized affricates /ts/ and /dʒ/ stemmed from (are historical counterparts of) the stops /t/ and /d/, probably, due to the intense palatalization of the latter (cf. Wexler 1977).
3. The analysis

The Byelorussian phonemes are classified non-traditionally in Table 2 according to the degrees of stricture (Saussure 1916 [2001]) and airflow: from the complete constriction to the maximal aperture, on the one hand, and from the complete stoppage of airflow to maximally free airflow, on the other hand (Davis 1984 [1987]).

Table 2. Phonetic Characterization of Byelorussian by Stricture and Airflow

<table>
<thead>
<tr>
<th>Stricture</th>
<th>Phonemes</th>
<th>Airflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>0°</td>
<td>p, t, k, b, d</td>
<td>0°: stopped</td>
</tr>
<tr>
<td></td>
<td>p', k', b'</td>
<td></td>
</tr>
<tr>
<td>0°</td>
<td>m, n</td>
<td>2°: nonturbulent</td>
</tr>
<tr>
<td></td>
<td>m', n'</td>
<td></td>
</tr>
<tr>
<td>0°-1°</td>
<td>ts, dz, ts' , dz'</td>
<td>0°-1°: stopped, then turbulent</td>
</tr>
<tr>
<td>1/2°</td>
<td>r</td>
<td>periodic</td>
</tr>
<tr>
<td>1°</td>
<td>f, s, š, x, v, z, ž, h</td>
<td>1°: turbulent</td>
</tr>
<tr>
<td></td>
<td>f', s', š', x', v', z', h'</td>
<td></td>
</tr>
<tr>
<td>2°</td>
<td>l, j, i, w, u</td>
<td>2°: nonturbulent, potentially turbulent</td>
</tr>
<tr>
<td></td>
<td>l'</td>
<td></td>
</tr>
<tr>
<td>3°-4°</td>
<td>o, ε</td>
<td>3°-4°: nonturbulent</td>
</tr>
<tr>
<td>5°</td>
<td>a</td>
<td>5°: nonturbulent</td>
</tr>
</tbody>
</table>

In Table 2., all the phonemes of Byelorussian are classified into seven categories.

1. Twelve phonemes of constriction are produced by complete stricture that can be done either by complete stoppage of airflow – stops or occlusives /p, t, k, b, d, p', k', b'/ in traditional phonetic terminology, or by closing the oral cavity with the uvula and creating the non-turbulent airflow through the nasal cavity – “nasals” /m, n, m', n'/. The nasals, although being phonemes of constriction, allow for the free non-turbulent airflow like phonemes of aperture (vowels, in traditional terminology), therefore, they are lower in the hierarchy than the stops.

2. Five phonemes, affricates, /ts, tj, dz, ts', dz'/, “have two degrees of stricture each, first complete stricture, then a transition to turbulent airflow” (Davis 1984 [1987]: 32).

3. The trilled /r/ is produced by periodic vibration and flapping of the apex against the alveolar ridge. This variability of /r/ represents the degrees of stricture varying around one half.
4. Fourteen phonemes, /f, s, ʃ, x, v, ʒ, h, ʃ, s, ʃ, ʃ, x, v, z, ţ, h/, are characterized by incomplete constriction of the oral cavity that creates turbulent airflow – fricatives in traditional phonetic terminology.

5. Besides the nasals, four other phonemes of constriction, like the phonemes of aperture, have non-turbulent airflow: /l, l, j, w/. In the case of the laterals /l, l/, the apex touches the upper teeth (the “hard” /l/) or the soft palate (the “soft” /l/) leaving a space on both sides of the tongue in the oral cavity for non-turbulent airflow. Although the semivowels /j, w/ hinder the air stream at some point, they have, like vowels, a high degree of aperture allowing for non-turbulent airflow. At the same time, the degree of stricture of two phonemes of aperture (vowels) /i/ and /u/ is low, making them similar to the laterals or the semivowels /j/ and /w/.

6. The two phonemes of aperture /o, ɛ/ are grouped together under the same degree of stricture between three and four replacing the traditional categories of close and open applied to the vowels of Byelorussian.

7. The highest degree of aperture is attributed to the neutral low central /a/.

The fronted versions of /u/, /ε/, /o/ and /a/ (shown in parentheses in Table 1.) are the variants (allophones) in the position after /j/ reflecting the orthographic distinctions between ‘у’ [u] and ’ю’ [ju], between ‘э’ [ɛ] and ’е’ [je], between ‘о’ [o] and ’ё’ [jo], and between ‘а’ [a] and ’я’ [ja].

As previously mentioned, Table 1. gives the traditional inventory of phonemes of Byelorussian, which are characterized by place and manner of articulation. However, it is not enough to know the descriptive characteristics of phonemes, namely, what is their “passive receptor (place of articulation)” to understand and account for their distribution in language. It is more important to know how a sound is articulated, i.e. what is “the articulator rather than the place of articulation” (Diver 1979: 174) that contributes to produce a sound. Table 3, following Davis (1984 [1987]), presents the recategorized phonemic inventory of Byelorussian based on active articulators, degrees of stricture and airflow.
Table 3. Phonemes of Byelorussian Characterized by Articulators, Stricture and Airflow*

<table>
<thead>
<tr>
<th>Oral Articulator(s)</th>
<th>Co-articulators</th>
<th>Lip</th>
<th>Apex</th>
<th>A-Dor</th>
<th>Lip + A-Dor</th>
<th>Apex + A-Dor</th>
<th>P-Dor</th>
<th>A + P-Dor</th>
<th>Lip + P-Dor</th>
<th>Gl</th>
<th>Gl + A-Dor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stricture/Airflow</td>
<td>0</td>
<td>L</td>
<td>V</td>
<td>0</td>
<td>L</td>
<td>V</td>
<td>0</td>
<td>L</td>
<td>V</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>p</td>
<td>b</td>
<td>t</td>
<td>d</td>
<td>p</td>
<td>b</td>
<td>k</td>
<td>k</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>ts</td>
<td>tf</td>
<td>dż</td>
<td>ts</td>
<td>dż</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/2</td>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>f</td>
<td>v</td>
<td>s</td>
<td>z</td>
<td>f</td>
<td>ž</td>
<td>f</td>
<td>v</td>
<td>s</td>
<td>z</td>
<td>x</td>
</tr>
<tr>
<td>2</td>
<td>l</td>
<td>j</td>
<td>i</td>
<td></td>
<td>l</td>
<td>i</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>3</td>
<td></td>
<td>ζ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td></td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* “Lip” refers to the lips, “A-Dor” refers to the anterodorsum, “P-Dor” to the posterodorsum and “Gl” refers to the glottis. “0” are the sounds produced without participation of any additional articulator, “L” are the sounds produced with excitation of larynx (voicing), and “V” are the sounds produced with excitation of velum (nasality).
Table 3. demonstrates that the degrees of stricture and airflow alone are not enough to distinguish between the phonemes of a language. The tongue and the lower lip are two basic active articulators which allow for maximal phonological variation. Following Davis (1984 [1987]: 48), I have divided the tongue into three regions – the apex (the tip of the tongue), the anterodorsum (the blade), and the posterodorsum (the back) – according to which part of the tongue has been exploited to produce distinctive phonemes. The larynx (L) and the nasal cavity, with the velum (V) as articulator, create two additional distinctive features: voicing and nasality, respectively. (0) represents the absence of these features. Table 3. manifests this same holistic approach by classifying the Byelorussian phonemes given in Table 2. Just as with phonemes of constriction, the characteristics of phonemes of aperture are based on the position of the active articulators used to produce distinctive vowels. Thus, the traditional features of front versus back vowels indicate the relative position of the anterodorsum or the posterodorsum, and open versus closed vowels refer to different degrees of aperture resulting from the position and the movement of the tongue and lips.

4. Phonemes of constriction in Byelorussian

4.1 Stable versus mobile: Likes are favored

In his examination of the skewing of word-initial consonant clusters composed of stops and fricatives and either /t/ or /l/ in English monosyllabic words, Diver (1979) raises the problem of the non-occurrence of the /tl/, /dl/, and /sr/ clusters. The results indicate that stops or occlusives favor clusters with /t/ over /l/, and the absence of the /tl/ and /dl/ clusters are the manifestation of the maximal disfavoring of /l/. On the other hand, fricatives prefer to cluster with /l/ over /t/, and the absence of the /sr/ cluster is the manifestation of the maximal disfavoring of /t/.

Since the communication and human factors underlie Diver’s analysis and are common to all human languages, I will now apply Diver’s methods to examine whether the same favorings are also present in Byelorussian. The preferences for the skewing of word-initial consonant clusters composed of stops and fricatives whose second member is either /t/ or /l/ in monosyllabic words in Byelorussian are presented in Table 4.
Chapter 5. Phonology as human behavior: Byelorussian

Table 4. Stable versus Mobile: Likes Are Favored; Differences Are Disfavored

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>b</th>
<th>t</th>
<th>d</th>
<th>k</th>
<th>f</th>
<th>v</th>
<th>s</th>
<th>z</th>
<th>ž</th>
<th>x</th>
<th>h</th>
<th>Totals</th>
<th>Cross Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>M + r:</td>
<td>r</td>
<td>7</td>
<td>11</td>
<td>7</td>
<td>9</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46</td>
</tr>
<tr>
<td>M + l:</td>
<td>1</td>
<td>13</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>S + r:</td>
<td>r</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>S + l:</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>Drop</td>
<td>-6</td>
<td>9</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>-7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data in Table 4 clearly show that all the stops (occlusives), except for /p/, favor the trilled /r/ while all the fricatives (spirants), except for /h/, prefer the lateral /l/. The various traditional classifications and categorizations based on the concept of manner of articulation do not explain this skewing. Diver, on the other hand, proposes two alternative distinctive features based on the kinds of gestures required to articulate sounds. The stops are produced by closure and release of the active articulators and are, therefore, mobile sounds, while the fricatives are produced by holding the active articulators in a steady position and are, therefore, stable sounds.

*Stable* indicates that the articulatory organ employed in the production of the sound is relatively stationary during excitation of the resonant cavity. … Mobile indicates the opposite: the articulator is necessary in motion during sound production. …

Returning now to the skewings, in terms of stable versus mobile, we can make the single statement that combinations of sames are favored and combinations of differents are disfavored. (Diver 1979: 171)

The exceptional disfavoring of /pr/ clusters found in Byelorussian has also been found for other Slavic languages: Ukrainian and Old Church Slavonic as well as for Swedish, Yiddish, Catalan, Irish and Esperanto (cf. Tobin 2002). Tobin (2002: 210–211, 228–229) accounts for the apparent exception with the voiceless mobile phoneme /p/ by articulatory and acoustic hypotheses.

If (explosive) mobile phonemes favor mobile (trilled) r, then that favoring might be weakest for the least explosive of the mobile phonemes, which would be p, since it has the largest oral cavity, hence the weakest air pressure for explosion. (Tobin 2002: 228)

This leaves a question why it does not take place with /b/ that is weaker in explosion than /p/. This question remains open for a further study.

At the same time, Tobin, following Davis (1984 [1987]), remarks that the disfavoring of the initial /pr/ clusters should be compensated by the favoring of the initial /kr/ clusters because /k/ “has the smallest oral cavity and the strongest pres-
sure for explosion” (Tobin 2002: 210). Davis (1984 [1987]: 69–72) says the following about the greater explosive energy that is inherent in /k/:

/k/ is advantaged (+) over /t/ (0) for Explosion, having a smaller super-laryngeal volume behind the constriction, that is, between the larynx and the dorsum-to-velum contact, and hence a greater potential energy. (Davis 1984 [1987]: 70)

Together with the disfavoring of the initial /pr/ clusters, Byelorussian appears to display the favoring of the initial /kr/ clusters, as one can see from the data in Table 4. In addition, an analysis of dictionaries shows that among all the words containing /pr/ and /pl/ clusters, the words with /pr/ clusters are strongly favored over those with /pl/ clusters.

We also found two apparent exceptions in the consonant clusters beginning with stables /v/ and /h/. As can be seen in Table 4, voiced stable /v/ is the only fricative that does not appear in initial clusters. This non-occurrence may be explained by the history of Byelorussian. It is believed that labial fricative /v/ developed from mobile /w/ in the Slavic and Romance languages as well as in Hebrew (cf. Tobin 2002: 211, 229).

The particular problem of the interrelationships between labial occlusives and fricatives in general and the replacement of /w/ by /v/ relevant to our discussion here is discussed in greater length from the point of view of markedness and the asymmetry of phonological systems in Gamkrelidze (1978: 29–33) and from the point of view of acoustic phonetics and spectography in Pulgram (1959). (Tobin 2002: 211)

In initial preconsonantal position in Byelorussian, mobile /w/ was not replaced by the stable /v/, as it was in Russian. Thus, in Russian, узлом [uzlom] became взлом [vzlom] 'break-in', while in Byelorussian, this change did not take place. The rise of the initial /v/ from /w/ occurred only in prevocalic positions so that in Contemporary Byelorussian, the initial /v/ is used mostly as a prothetic prevocalic consonant (e.g. the Byelorussian во́друк [vodhuk] is о́тзвук [otzvuk] 'echo' in Russian). The fact that the voiceless stable /h/ favors mobile /t/ may be also accounted for by diachronic changes. In Byelorussian, stable glottal /h/ derived from velar mobile /g/ by the spirantization of the latter that had already taken place in the pre-Byelorussian pre-literary period. The synchronic favoring of /t/ may reflect the older correlation between posterodorsal mobile /g/ and mobile apical /r/.

The data in Table 4 show that the same principle that similar gestures (likes) are favored and different gestures (differents) are disfavored, which worked for English (Diver 1979), Italian and Latin (Davis 1987), Hebrew and Yiddish (Tobin 1997) and 40 other languages (Tobin 2000), is relevant for Byelorussian as well. One can see that there are 79 instances of likes, obtained by summing up all the
instances of similar gestures of mobility and stability (Mobile + Mobile, Stable + Stable), versus 52 instances of differents (Mobile + Stable, Stable + Mobile).

4.2 Maximal constriction is favored

Diver (1979), Davis (1984 [1987]), and Tobin (1997, 2002) also contrasted the differences between the frequencies of favored and disfavored distributions of the initial clusters whose second phoneme is either /r/ or /l/. Table 5 shows the differences for Byelorussian under the rubric of “Drop”. It is significant to notice that the average amount of drop in frequency for the mobile phonemes is greater than that for the stable phonemes. The difference is due to the nature of the mobile and stable sounds. Mobile phonemes require complete closure and release of the muscles of the active articulators, i.e. maximal constriction is followed by their aperture. By contrast, stable phonemes require greater effort to maintain a single degree of stricture by holding the active articulators in a steady position. In this connection, Davis (1984 [1987]: 50–52) proposes the principle of favoring of maximal constriction. The data of Table 5 collected from all initial consonant clusters in Byelorussian involve the difference in frequency depending on the change in stricture.

Table 5. Maximal Constriction Is Favored in Clusters

| 0° stricture | p | 24 | b | 13 | k | 28 |
| 1° stricture | f | 8  | v | 0  | x | 14 |

One may see that in Table 5, the mobile phonemes with complete constriction (0°) are more frequent in initial position than their stable turbulent (1°) counterparts.

Though mobile obstruents are easier to produce and are acquired earlier than stable ones, in Byelorussian, as in most languages, there are more stable obstruents than mobile ones. This can be explained by the human factor. As previously stated, mobile obstruents are produced by complete stoppage of airflow, while stable ones allow for turbulent airflow. Therefore, physiological limitations, “imposed” to mobile obstruents, do not always allow us to produce a complete constriction in the similar phonetic environments where stable obstruents appear.

4.3 Initial visible articulations are favored

Data in Diver (1979: 179) show that in word-initial position there is a slight preference in frequency for labial phonemes of constriction. Davis (1984 [1987]: 44–47) refers to the favoring of visible articulators in initial position as Visibility. Visible phonemes like the labials can be seen as well as heard, appealing to both our senses
of vision and hearing. These phonemes provide a greater number of clues, thus making communication easier and more efficient (cf. Tobin 1997: 44).

I will now examine whether the same favoring is also present in Byelorussian. The preference for labial phonemes of constriction in word-initial position in Byelorussian is presented in Table 6.

Table 6. Initial visible articulations are favored

<table>
<thead>
<tr>
<th>Visible Phonemes of Constriction</th>
<th>Initial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>2.5: 1</td>
<td></td>
</tr>
</tbody>
</table>

The data in Table 6 confirm that the same favoring of visible phonemes in word-initial position, which worked for English (Diver 1979), Italian and Latin (Davis 1984 [1987]), Hebrew and Yiddish (Tobin 1997), and other languages (Tobin 2000), is present in Byelorussian as well. The data show that visible phonemes are more than twice as prevalent in word-initial position as in word-final position.

4.4 Final apical articulations are favored

Another property pointed out by Diver is the adroitness of the apex, which appears to play an important role in the distribution of the apical phonemes. The apex is the most flexible, the most sensitive and the easiest to control of all the active articulators. Therefore, it is not surprising that of 36 phonemes of constriction, 14 (i.e. almost half) are articulated with the apex.

Since the apical sounds are the easiest to produce, speakers may use them in those positions in a word where minimal effort may be applied without detriment to communication. In word-final position, the importance of communication is decreased both for the encoder and for the decoder because most of the information necessary for maintaining distinctiveness has been previously given, in the beginning and in the middle of the word. Diver (1979), Davis (1984 [1987]), and Tobin (1997) showed that the apical sounds are significantly preferred in word-final position in English, Italian and Hebrew, respectively. I will now examine whether the same tendency is relevant to Byelorussian. The preference of apical phonemes in word-final position in Byelorussian is observed in Table 7.
Table 7. Final Apical Articulations Are Favored

<table>
<thead>
<tr>
<th>Apical Phonemes of Constriction</th>
<th>Initial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>302</td>
<td></td>
<td>368</td>
</tr>
<tr>
<td>Ratio 1.22: 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data in Table 7 validate the favoring of apical phonemes in word-final position.

4.5 Additional articulators are disfavored

4.5.1 One articulator (voiceless) is favored; two articulators (voiced) are disfavored

The degree of muscle control required to articulate sounds may influence the distribution of voiced and voiceless obstruents. Since voicing is the simultaneous excitation of the oral tract and the glottis, it is natural to suppose that the greater effort brought about by the addition of an articulator, i.e. the glottis, will lead to disfavor voiced sounds.

In the case of what is called a voiceless stop – $t$, for example – a single articulator both shapes and excites the resonant cavity; for $d$ on the other hand, an additional articulator – the glottis – is invoked to provide additional excitation, thus changing the character of the acoustic product and adding another distinctive unit to the inventory. But coordination of the glottis with the apex, doing two things at once, requires more control, and the voiced stops suffer in frequency accordingly… (Diver 1979: 174–175)

Thus, Diver appeals to the human factor to explain the favorings and disfavorings found in the English data. The same explanation can account for all the voiced obstruents in Byelorussian. Table 8 presents the results of the skewing of initial consonant clusters in monosyllabic words in Byelorussian, depending on the number of the sets of articulators.

The data in Table 8 have been taken from the larger inventory of all initial consonant clusters in Byelorussian and show that the voiceless clusters (produced by one set of articulators) consistently outnumber the voiced clusters (employing both oral articulators and the larynx).
Table 8. Additional Articulators Are Disfavored. One Articulator (Voiceless) Is Favored; Two Articulators (Voiced) Are Disfavored

<table>
<thead>
<tr>
<th></th>
<th>Voiceless</th>
<th>Voiced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p 24</td>
<td>b 13</td>
</tr>
<tr>
<td></td>
<td>t* 16</td>
<td>d 13</td>
</tr>
<tr>
<td></td>
<td>s 69</td>
<td>z 42</td>
</tr>
<tr>
<td></td>
<td>f 22</td>
<td>ž 42</td>
</tr>
<tr>
<td></td>
<td>f 8</td>
<td>v 0</td>
</tr>
<tr>
<td></td>
<td>ts 7</td>
<td>dž 1</td>
</tr>
<tr>
<td></td>
<td>tf** 5</td>
<td>dž 1</td>
</tr>
</tbody>
</table>

* The difference in numbers of initial consonant clusters for monosyllabic words beginning with mobiles /t/ and /d/ was extremely low. Therefore, data were also collected from disyllabic words containing these consonant clusters. The data (/t/: /d/ = 60: 28) more strongly support this principle and reveal that the disyllabic words whose initial consonant clusters begin with the voiceless mobile /t/ are greater in number than the disyllabic words whose initial consonant clusters begin with the voiced mobile /d/.

** The difference in numbers of initial consonant clusters for monosyllabic words beginning with affricates /t∫/ and /dž/ was extremely low. Therefore, I examined all the words containing consonant clusters with these sounds. I found no more clusters with /dž/ versus six with /t∫/. Thus, the data (/t∫/: /dž/ = 11: 1) more strongly support the disfavoring of additional articulators.

4.5.2 One articulator in final position is favored

The principle of disfavoring sounds produced with two sets of muscles may influence not only their combinations with other consonants but also their distribution within the word itself. As previously stated, in word-final position, the importance of communication is decreased. Therefore, one might expect that sounds requiring more effort and a greater degree of control (i.e., the voiced sounds with two articulators versus the voiceless sounds with one articulator) will be disfavored in word-final position where the burden of communication is lower. Diver states the following on this matter:

It is entirely understandable that if different gestures present different degrees of difficulty, the more difficult gestures will be disfavored where the motivation for maintaining distinctiveness is lessened, as in final position. (Diver 1979: 179)

Table 9 presents the results of the distribution of obstruents in word-initial and word-final positions in Byelorussian, depending on the number of articulators.

Table 9. Additional Articulators Are Disfavored. One Articulator in Final Position Is Favored

<table>
<thead>
<tr>
<th>Number of articulators</th>
<th>1 articulator</th>
<th>2 articulators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>378</td>
<td>163</td>
</tr>
<tr>
<td>Final</td>
<td>365</td>
<td>0</td>
</tr>
</tbody>
</table>
The data in Table 9 show that the frequency of voiceless and voiced obstruents in word-initial position (high communicative load) (378 to 163) descends to zero in word-final position, where the need for distinctiveness is diminished. Here, as in Russian and in German, the communicative benefit of making additional effort to coordinate two articulators drops and leads to a complete devoicing of final consonants (365 to 0).

At the same time, the voiced sounds (including voiced semivowels, liquids, approximants as well as the vowels) in Byelorussian, as in other languages, exceed in number the voiceless ones, even though the voiceless sounds are easier to make. This circumstance does not contradict PHB and can be accounted for by the communication factor. Though the voiced sounds require “an investment in greater precision of control” (Diver 1995: 65), they allow us to increase the basic inventory. Consider Tobin (1997) on this matter:

Voiced sounds provide more acoustic and perceptual information – they provide resonance and have acoustic formants that provide the primary auditory and communicative information most necessary for speech perception. Therefore, the need for maximum communication justifies the extra effort necessary for making voiced sounds. (Tobin 1997: 316, n. 7)

4.5.3 Coordination of different activities (voicing and nasality) is disfavored
The same principle of disfavoring of additional articulators is also true for nasals, because the production of nasals involves the simultaneous coordination of three sets of articulators: the active articulator (the lips or the apex), the glottis and the uvula. Therefore, one might expect that clusters with nasals will be even more disfavored in Byelorussian than clusters with voiced sounds. Table 10 shows the results of the skewing of initial consonant clusters for labials and apicals, depending on the number of articulators.

<table>
<thead>
<tr>
<th>Articulators</th>
<th>Labial</th>
<th>Apical</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 articulator</td>
<td>p</td>
<td>t</td>
</tr>
<tr>
<td>2 articulators</td>
<td>b</td>
<td>d</td>
</tr>
<tr>
<td>3 articulators</td>
<td>m</td>
<td>n</td>
</tr>
</tbody>
</table>

The data of all initial consonant clusters in Table 10 indicate a decrease in number for each case of the addition of one articulator (voicing [+1] and then nasality [+2]).
Palatalization is disfavored

As previously mentioned, Byelorussian has in its inventory palatalized (“soft”) counterparts of “hard” phonemes as a subsystem within the larger system of phonemes of constriction. Palatalized phonemes are produced by excitation of an additional active articulator (the anterodorsum) in the direction of a passive receptor (the soft palate). Such a process requires a greater effort in a way similar to voicing and nasalization and, therefore, should be disfavored. Table 11 shows the results of the comparison between different consonants in word-initial (Initial) and word-final (Final) positions, depending on the addition of palatalization which requires the excitation of another set of muscles.

The data in Table 11 indicate a strong disfavoring of most palatalized (“soft”) phonemes both in word-initial and especially in word-final positions. For example, for labial mobile /p/, there is a drop of 35 (from 42 to 7), i.e. 83%, when its palatalized counterpart is used in word-initial position and a drop to zero, i.e. 100%, in word-final position. For labial nasal /m/, there is a drop of 11 (from 25 to 14), i.e. 44%, in word-initial position and a drop to zero, i.e. 100%, in word-final position.

At the same time, the distributions of apical stables /s, z/ and /n/ differ from the distributions of other phonemes. These sounds are the only ones whose palatalized counterparts appear in word-final position. I appeal to the concept of active articulators to explain the distribution of palatalized /s, z/ and /n/. I believe that palatalized apicals are less disfavored than other palatalized phonemes because of the adroitness of the apex which has been previously discussed. The preference by 1 for /n/ over /n/ is rather idiosyncratic.
Table 11. Additional Articulators Are Disfavored. Palatalization Is Disfavored

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Final</th>
<th>Initial</th>
<th>Final</th>
<th>Initial</th>
<th>Final</th>
<th>Initial</th>
<th>Final</th>
<th>Initial</th>
<th>Final</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Hard”</td>
<td>42</td>
<td>12</td>
<td>33</td>
<td>21</td>
<td>25</td>
<td>20</td>
<td>5</td>
<td>31</td>
<td>82</td>
<td>27</td>
<td>48</td>
</tr>
<tr>
<td>“Soft”</td>
<td>7</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>6</td>
<td>12</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>% of Drop of “Hard”</td>
<td>83%</td>
<td>100%</td>
<td>82%</td>
<td>100%</td>
<td>44%</td>
<td>100%</td>
<td>-20%</td>
<td>61%</td>
<td>95%</td>
<td>81%</td>
<td>94%</td>
</tr>
</tbody>
</table>
4.6 Additional gesture is disfavored: Transitions from one distinct constriction to another within a single phoneme are disfavored

Four phonemes in Byelorussian /ts, dz, tj, dž/ have the characteristic of transition from complete constriction (0°) to slight aperture (1°) (affricates, in traditional terminology). The performing of two consecutive constrictions in one gesture requires more effort than one constriction and, therefore, will be disfavored. Davis (1984 [1987]: 58–59) calls this factor T, for “Transition”: “Transitions from one distinct constriction to another within a single phoneme are disfavored.” The data are in Table 12.

<table>
<thead>
<tr>
<th></th>
<th>t 116</th>
<th>d 66</th>
<th>f 54</th>
<th>ž 22</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 constriction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 constrictions</td>
<td>ts 90</td>
<td>dz 14</td>
<td>tf 34</td>
<td>dž 5</td>
</tr>
</tbody>
</table>

The data of Table 12, taken from all the words in Byelorussian with phonemes of transition and their counterparts with complete constriction, indicate a clear-cut disfavoring in the distribution of the former.

4.7 Reuse of the same musculature is disfavored

Communication depends on a physiological process of the production of sounds through a sequence of articulatory gestures. Combinations of gestures that are easier to articulate are also easier to learn to control. Such combinations will be preferred over combinations that are more difficult to produce and, hence, are more difficult to learn to control. Diver (1979: 176) compares and contrasts the influence of different disfavorings on the frequency of phonemes: “we must take into account how many disfavorings are present in a given instance: … two disfavorings have greater consequence than one.”

As stated by Diver, the disfavoring of different gestures (Mobile + Stable, Stable + Mobile) over similar gestures (Mobile + Mobile, Stable + Stable), discussed earlier, has the least impact on frequency. The disfavoring of additional articulators (voicing and nasality) over one articulator (voiceless sounds), also discussed earlier, causes a greater effort in the production of phonemes and, therefore, affects frequency more strongly. Diver explains the extreme disfavorings of initial /tl/, /dl/ and /sr/ clusters in English by the reuse of the same muscles in the transition from mobile to stable.
With \( tl^- \), for example, we have a violent motion, the explosion that must immediately be quelled and brought under control so that a gesture of a very different kind can be made. With \( tr^- \), on the other hand, the violence of the motion can continue unchecked; it is merely transmuted – “steered” perhaps is the term – into another gesture equally violent, and the control problem is greatly lessened. Where we begin with a stable gesture, \( s \) or \( f \), we have the same problem of control in reverse. (Diver 1979: 173)

The same disfavoring manifests also in Byelorussian. Table 4 has shown a tendency to avoid reusing the same articulator. However, because the number of consonant clusters taken from monosyllabic words in Byelorussian was low, I took data from disyllabic words containing the same consonant clusters. The data are found in Table 13.

**Table 13.** Reuse of the Same Musculature Is Disfavored. Non-random Distribution of Disyllabic Words

<table>
<thead>
<tr>
<th></th>
<th>( t )</th>
<th>( d )</th>
<th>( s )</th>
<th>( f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M + r: )</td>
<td>( r )</td>
<td>50</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>( M + l: )</td>
<td>( l )</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>( S + r: )</td>
<td>( r )</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>( S + l: )</td>
<td>( l )</td>
<td>25</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

The data in Table 13 reveal that the disfavoring of the reuse of the same musculature is even more strongly supported for disyllabic words.

Diver (1979: 176–178) illustrates how the piling up of disfavorings influences the frequency of words in English. He discusses the distributions of /\( l \)/ and nasals in monosyllabic words containing consonant clusters of the following types:

- **CVC** 189 \( slip \)  \( CNVC \) 60 \( snip \)
- **CCVl** 40 \( spill \)  \( CCVN \) 98 \( spin \)
- **CIVl** 1 \( flail \)  \( CNVN \) 0 \( smin \)

Diver shows the drop in the number of words existing in English when the same articulator is reused in word-final position “after appearing in an initial cluster” (Diver 1979: 176–177). Diver pronounces on this matter as follows:

First, with \( l \), there are 189 words of the type \( slip \), with \( l \) in the initial cluster, and 40 words of the type \( spill \), with \( l \) in final position, in the presence of an initial cluster. But there is only one word, \( flail \) with \( l \) in both the initial cluster and final position.

So also with nasals … (Diver 1979: 177)
The principle of avoidance of the reuse of the same articulator has also been found for Italian (Davis 1984 [1987]: 42), Hebrew and Yiddish (Tobin 1997: 109–111, 116).

Parallel data found in Byelorussian appear in Table 14 My table has three columns instead of two, as in Diver, because I have examined the distributions of /r/ along with those of /l/ and nasals in monosyllabic words with consonant clusters. It has four rows instead of three, as in Diver, because I have examined the distributions of these phonemes in the initial cluster, in final position in the presence of an initial cluster, in the final cluster and in both the initial cluster and final position.

Table 14. Reuse of the Same Musculature Is Disfavored. Reuse of the Same Articulator Both in Initial and Final Positions Is Disfavored

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVC</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>CCVI</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>CVCI</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CVNI</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

With /l/, there are 48 words of the type [khlam] with /l/ in the initial cluster, 10 words of the type [stol], with /l/ in final position in the presence of an initial cluster and only 2 words of the type [tsykl], with /l/ in the final cluster. Moreover, there are no words of the type [sllal], with /l/ in both the initial cluster and final position. Similar patterns hold for /r/ and /n/.

Diver (1979: 176–178) also examined monosyllabic words of simple CVC type and found that the disfavoring of the reuse of the same articulator holds even there, where clustering is not an additional factor. In Table 15, I consider the combinations of the initial and final labial, apical and dorsal stops, as they appear in Byelorussian in monosyllabic words of the simple CVC type, as in the word [bok].

Table 15. Reuse of the Same Musculature Is Disfavored. Observed Combinations of Initial and Final Stops

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>D</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>
I have found 21 monosyllabic words both beginning and ending with stops. Nine of the 21 begin with labials /p, b/, seven with apicals /t, d/, and five with dorsal /k/. Three words end with labials, eleven with apicals, and seven with dorsals. If on this basis one calculates a symmetrical distribution in the combination of initials and finals, assuming that the reuse of the same articulators does not influence the distribution of phonemes, one gets the calculated values for the co-occurrences of initial and final stops. For example, the calculated value for the words of the type [bok] is obtained by taking the proportion of initial labials (9/21) and multiplying it by the proportion of final dorsals (7/21). The product is multiplied by 21.

\[(9/21) \times (7/21) \times 21 = 3\]

The obtained result (3) is the calculated value for labials plus dorsals (L+D) that could be expected to occur if all initials and finals combined freely with each other. Table 16 presents both the observed and the calculated results of the co-occurrence of initial and final stops in Byelorussian.

### Table 16. Reuse of the Same Musculature Is Disfavored. Reuse of Articulator at Beginning and End of Word: Comparison between the Calculated and the Observed Co-occurrences

<table>
<thead>
<tr>
<th></th>
<th>L + A</th>
<th>L + D</th>
<th>A + L</th>
<th>A + D</th>
<th>D + L</th>
<th>D + A</th>
<th>L + L</th>
<th>A + A</th>
<th>D + D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculated</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Observed</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Difference</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>-3</td>
<td>-2</td>
</tr>
</tbody>
</table>

* The calculated value is a rounded number; therefore, the summing up of the calculated outcomes may differ from the observed results.

The left section of the calculated values displayed in Table 16. shows the combinations of different articulators at the beginning and the end of the word, of the types [bot] (L+A), [bok] (L+D), [dub] (A+L), [tok] (A+D), [kab] (D+L), and [kot] (D+A). The combinations of the same articulators, of the types [bob] (L+L), [tut] (A+A), [kok] (D+D), are presented in the right section of Table 16.

Byelorussian shows a tendency to avoid reusing the same articulator. The left side of the table shows that, overall, different articulators are used as much or more than would be predicted for a symmetrical distribution (differences of 0, +1, and +2). The right side shows the reverse tendency (-1, -3, -2). That is, almost all combinations with different articulators increase a little, and all those with the same articulators drop.
5. **Summary and conclusions**

In this paper, PHB has been applied to Byelorussian. Following Diver (1979), Davis (1984 [1987]), and Tobin (1997), I have replaced some traditional phonetic and phonological categories with alternative ones which are more directly oriented to human physiology, perception, cognition and behavior:

- the active articulators used to produce sounds versus the traditional concept of *place of articulation* indicating the passive receptors;
- the kinds of gestures producing sounds and changing the airflow (mobile/stable) versus the traditional concept of *manner of articulation*;
- the number of sets of articulators required to produce sounds versus the traditional concepts of *voiceless/voiced/nasal*;
- the single hierarchy of sounds based on relative degrees of stricture and airflow and types of airflow versus the traditional categories of consonants and vowels and the various traditional features related to manners of articulation.

The distribution of phonemes in Byelorussian can be accounted for by the “mini-max” principle of “a conflict between the communication and the human factors in the search for maximum communication with minimal effort” (Tobin 1997: 47). Throughout this paper, a correlation has been observed between the effort applied by speakers to control the active articulators involved in the production of phonemes and the observed favorings and disfavorings of these phonemes in various phonotactic distributions.

The following principles concerning phonemes of constriction have been validated:

1. consonant clusters composed of similar gestures are favored over clusters composed of different gestures;
2. among constrictions, maximal constriction is favored;
3. visible phonemes are favored in word-initial position;
4. apical phonemes are favored in general and particularly in word-final position;
5. additional articulators are disfavored;
6. transitions from one distinct constriction to another within a single phoneme are disfavored;
7. reuse of the same articulators is disfavored.
Chapter 5. Phonology as human behavior: Byelorussian

Notes

1. I gratefully acknowledge Professor Yishai Tobin whose support, guidance and advice helped me in writing this paper. I would also like to thank anonymous reviewers for their comments and suggestions. Despite the meticulous attention of all of them to details, I will claim possible errors of fact, judgment and style for myself only.

2. Cf. Davis (this volume) on the phoneme as theoretical unit.

3. Diver notes that, because of the communicative factor, apicals are not particularly favored at the beginning of a word, as they are at the end.

References


**Dictionaries**


Phonology as human behavior

The case of Peninsular Spanish

Adriaan Dekker and Bob de Jonge
University of Groningen

The authors analyze the distribution in the lexicon of nine Peninsular Spanish consonants (/p, t, k, b, d, g, f, θ, χ/) within the framework of Phonology as Human Behavior with respect to two hypothesised factors: Complexity of Articulation and Visibility. In general, the observed distribution of the uncombined consonants is according to the hypotheses tested. However, one member, /k/, occurs more frequently than would be expected. The authors show that /k/ belongs to a particular subset of consonants in Spanish, which might explain a relative favoring, but they also observe that /k/ appears to be more frequent in other languages as well. It appears that another factor, the Size of Cavity, might play a decisive role.

1. Introduction

In this paper, the theory of Phonology as Human Behaviour (PHB) (Diver 1975; Tobin 1997) will be applied to a corpus of Peninsular Spanish word stems. In particular, we analyze the distribution in the lexicon of nine consonants (/p, t, k, b, d, g, f, θ, χ/) with respect to two hypothesised factors: Complexity of Articulation and Visibility. In the analysis, certain data appear to contradict the general hypothesis. We will try to resolve this problem by comparing systematically both minimal pairs and minimal triplets, and by appealing to characteristics of sets and subsystems of phonemes in the language. In the end, a general conclusion will be drawn.
2. The data and procedure

The data were gathered by taking monosyllabic word stems from a Spanish dictionary (Heras Fernández and Rodríguez Alonso 1993). The analysis of the complete dictionary rendered a total of 1986 monosyllabic stems, of which 291 begin with a consonant cluster (CCV-), and 1695 with a single consonant (CV-). Of this last, unclustered, group, only the initial sets /p, t, k/ (unvoiced stops), /b, d, g/ (voiced stops), and /f, θ, χ/ (unvoiced fricatives), were used for the present paper for reasons of inherent comparability, rendering a total of 1030 cases. The fact that each group consists of three comparable consonants permits us to compare frequencies of occurrence of each group in total, as well as of each member individually, compared with its counterparts in one of the other groups. This group of data was subjected to hypotheses established in earlier studies of PHB, such as Diver (1975), Davis (1984 [1987]), and Hameed (1999). Below, we will discuss the results for Complexity of Articulation and the Visibility Hypothesis. Then, we will present an additional explanation in order to resolve the distributional problem that rises from these two hypotheses taken together.

We have to bear in mind that we are dealing here not with minimal pairs, but with sets of triplets. First, we will make comparisons on the basis of minimal pairs, which enables us to make predictions for the distribution of one group, say, /p, t, k/, with each of the other ones, /b, d, g/ and /f, θ, χ/, respectively. Then, a comparison among three groups can be made, because each group has an expected relationship to each of the other groups.

3. Complexity of articulation

Relevant to this study, Complexity of Articulation takes two forms: the number of articulators required for the production of a phoneme, and the relative degree of control required to execute that gesture.

Alarcos Llorach defines voiced consonants as being a sum of the voiceless counterpart and voicing:

\[
\text{[…] el rasgo se debe a un generador armónico suplementario al generador consonante normal de las consonantes; se suma la sonoridad originada en las cuerdas vocales.} \]

(Alarcos Llorach 1964: 71)

‘[… ] the feature has its origin in a harmonic generator, supplementary to the normal generator of the consonants; voicing, produced by the vocal strings is added to them.’ (emphasis added, AD and BdJ)
Diver speaks in terms of an “additional articulator” in order to explain the greater effort that is required to produce voiced stops:

In the case of what is called a voiceless stop, $t$ for example, a single articulator both shapes and excites the resonant cavity; for $d$ on the other hand, an additional articulator, the glottis, is invoked to provide additional excitation, thus changing the character of the acoustic product and adding another distinctive unit to the inventory. (Diver 1975: 47)

The Additional Articulator Hypothesis supposes that consonants that require more than one articulatory gesture (i.e., oral plus voicing) are more difficult to produce than simpler ones and will therefore occur less frequently than consonants that require only a single articulatory gesture. Davis describes this principle as follows:

…the coordination of different activities is known to be a problem in human behavior in general. Diver uses the analogy of simultaneously patting the head and rubbing the tummy, or vice versa. One also thinks of the accomplished pianist who sets out to learn to play the organ and spends several frustrating week trying to coordinate hands and feet. Of course, the difficulty involved in activating the larynx is much less severe, and people pronounce $b$’s and even $m$’s with no appreciable difficulty; yet in the long life of language, the disfavoring produced by the extra articulator has its effect on frequency. (Davis 1984 [1987]: 41)

The Explosion Hypothesis (Davis 1984 [1987]: 52) predicts that stops will be more frequent in initial position than their fricative counterparts. In order to produce a fricative, a speaker first has to create pressure in the lungs, just as with plosives. But then, he or she has to free this pressure in a controlled manner, in contrast with plosives, where this release is uncontrolled. Therefore, the production of a fricative is more complex and will be likely to occur less frequently in comparison to the relatively easy unvoiced stops.

The first hypothesis to be tested in our corpus is therefore related to Complexity of Articulation. The expectation is that in initial position, the group of the unvoiced stops $/p, t, k/$ will occur most frequently, in view of the fact that the voiced stops $/b, d, g/$ have an additional articulator, the larynx and the group of unvoiced fricatives $/f, \theta, \chi/$ have a more difficult articulation. The results are shown in Table 1.
Table 1. Absolute and relative frequencies of different consonants in initial word position according to simplicity or complexity of articulator type

<table>
<thead>
<tr>
<th>Articulator type</th>
<th>#C</th>
<th>Observed frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple: unvoiced stops</td>
<td>/p, t, k/</td>
<td>492 (48%)</td>
</tr>
<tr>
<td>Complex: voiced stops</td>
<td>/b, d, g/</td>
<td>294 (28%)</td>
</tr>
<tr>
<td>Complex: unvoiced fricatives</td>
<td>/f, θ, χ/</td>
<td>244 (24%)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1030 (100%)</td>
</tr>
</tbody>
</table>

Table 1 confirms our expectations. The unvoiced stops, supposedly the consonants least difficult to produce, are observed more frequently than comparable voiced stops and comparable fricatives: unvoiced stops represent almost 48% of all cases vs. the other two: 28% and 24% respectively. If the degree of complexity were not to influence this distribution, equal percentages would be expected in all cases, i.e., 33.3% for each group. The latter two groups do not differ very much among each other. This may not be too surprising, since both sets require two basic actions from the speaker: in the case of the voiced stops, the creation of pressure and voicing; in the case of the fricatives the controlled release of air.

4. The Visibility Hypothesis

The next hypothesis to be tested in our corpus is the Visibility Hypothesis. This hypothesis supposes that the more visible a consonant is in initial position, the more effective its communicative impact, and therefore the more frequently speakers will tend to use it. Davis argues as follows:

Factor V [Visibility] demonstrates the importance of communication in phonology by positing a favoring for visibility in initial position of the word, just where most of the distinctiveness of the word is at stake: initially, the hearer knows nothing; once he knows that the word begins with, say, /p/, he is well on his way to comprehension, having eliminated a couple of dozen other possibilities. (Davis 1984 [1987]: 45)

After having reordered our group of consonants in three groups which differ in degree of visibility, what we expect to see is the following: the members of the group of consonants which are more visible (p, b, f) should occur most frequently, the group that is less visible (t, d, θ) less frequently, and the least visible ones (k, g, χ) least frequently. Table 2 gives the results of this test.
### Table 2. Absolute and relative frequency of different consonants in initial word position according to degree of visibility*

<table>
<thead>
<tr>
<th>Visibility</th>
<th>#C</th>
<th>Observed frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ visible</td>
<td>labials (p, b, f)</td>
<td>396 (39%)</td>
</tr>
<tr>
<td>± visible</td>
<td>apicals (t, d, θ)</td>
<td>311 (30%)</td>
</tr>
<tr>
<td>– visible</td>
<td>dorsals (k, g, χ)</td>
<td>323 (31%)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1030 (100%)</td>
</tr>
</tbody>
</table>

* The data presented in this paper do not represent a sample, and so these are descriptive, not inferential statistics. Therefore, statistical significance is not an issue and will therefore not be used in this article.

The results partially meet with our expectations: the labials are clearly more frequent than apicals and dorsals respectively (39% vs. 30% and 31%). But since apicals are slightly more visible than dorsals, we would also expect a higher frequency for the apicals, but such is not the case. The observed frequencies, although with a very small difference, even go counter to our expectations (precisely, 30% for apicals vs. 31% for dorsals).

### 5. The problem

We now examine more closely two questions: what is the explanation for the small difference between voiced stops and voiceless fricatives? And what is the origin of the deviation from the expected difference in percentages between apicals and dorsals? In order to answer these questions, we state the theoretical influence of these two factors in positive (+), intermediate (±) and negative (−) terms and indicate the combination of these terms in Figure 1.

#### Figure 1. Indication of positive and negative influences by two factors on production of individual consonants

<table>
<thead>
<tr>
<th>Articulator type</th>
<th>Simple: unvoiced stops +</th>
<th>Complex: voiced stops ±</th>
<th>Complex: fricatives –</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ visible</td>
<td>/p/: + +</td>
<td>/b/: + ±</td>
<td>/f/: + –</td>
</tr>
<tr>
<td>± visible</td>
<td>/t/: ± +</td>
<td>/d/: ± ±</td>
<td>/θ/: ± –</td>
</tr>
<tr>
<td>– visible</td>
<td>/k/: – +</td>
<td>/g/: – ±</td>
<td>/χ/: – –</td>
</tr>
</tbody>
</table>
Translated into numbers, this should mean that /p/ (with two pluses) is to occur most frequently of all consonants, then /b/ and /t/, then /k/, /d/ and /f/, and then numbers should drop from /g/ and /θ/ to /χ/ as the consonant with the lowest frequency of all.

To control this theoretical scale of frequencies, we will first calculate the expected distribution of the individual consonants of the investigated groups on the basis of the totals of the consonant groups, observed in Tables 1 and 2. This will give the hypothetical symmetrical distribution of the nine consonants if Complexity and Visibility were not factors in the distribution. In Table 3, the bottom row consists of the totals of Table 1, the right hand column of the totals of Table 2. The (calculated) expected frequency of, say, /p/, is found by multiplying its respective totals, 492 and 396. This result is divided by 1030. The outcome is 189.16, which gives a calculated expected frequency of 189 occurrences. This calculation was executed for all consonants in the table, with the data of Table 3 as a result.

Table 3. Expected absolute frequencies of individual consonants, on the basis of Tables 1 and 2

<table>
<thead>
<tr>
<th>#C</th>
<th>Expected frequency</th>
<th>#C</th>
<th>Expected frequency</th>
<th>#C</th>
<th>Expected frequency</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
<td>189</td>
<td>/b/</td>
<td>113</td>
<td>/f/</td>
<td>94</td>
<td>396</td>
</tr>
<tr>
<td>/t/</td>
<td>148</td>
<td>/d/</td>
<td>89</td>
<td>/θ/</td>
<td>74</td>
<td>311</td>
</tr>
<tr>
<td>/k/</td>
<td>155</td>
<td>/g/</td>
<td>92</td>
<td>/χ/</td>
<td>76</td>
<td>323</td>
</tr>
<tr>
<td>Total</td>
<td>492</td>
<td>Total</td>
<td>294</td>
<td>Total</td>
<td>244</td>
<td>1030</td>
</tr>
</tbody>
</table>

Comparing the results from Table 3 with Figure 1, it is clear that the calculated expected frequencies of Table 3 do not mirror the theoretical influence of the two factors under investigation, rendered in Figure 1. In Table 3 we see a slight rise in calculated frequency of /k, g, χ/ as opposed to /t, d, θ/, where we would predict a lower frequency on the basis of the former’s lesser inherent visibility from Figure 1; that is, across the board, dorsals are more frequent in our corpus than the hypotheses of Complexity and Visibility predict. Also, we see a strong fall in expected frequency from /k/, on the one hand, to both /g/ and /χ/, on the other, where we might expect, on the basis of Figure 1, to observe a gradual fall from /k/ to /g/ and /χ/ (This is assuming that, in Figure 1, “+”, “±”, and “–” are linearly relative).

Things get even more puzzling when we look at the observed frequencies of the different consonants: as Table 4 shows, these too do not reflect the theoretical expectations of Figure 1, but show an even greater preference for /k/ than in Table 3. Table 4a shows the differences of expected and observed frequencies in absolute and in relative figures.
It is clear that /k/ has the highest frequency of all consonants, higher even than the theoretically preferred plosives /t/ and /p/: 194 tokens of /k/ were observed, a deviation of 39 (25%) above the expected frequency of 155. The two biggest negative deviations from the expected frequencies, in relative terms, are /g/ and /χ/, respectively. They deviate negatively by 25% and 21%, respectively.

The general pattern of Table 4 tells us, however, that in fact the only deviant member of the system is /k/, since /p/ scores higher than /t/, but also higher than /b/. These results are in accordance with the hypotheses presented in Tables 1 and 2. Also the other members, except /k/, behave exactly like the theoretical prediction of Figure 1. In Table 5 we show the resemblance between the distribution of Table 4 and Figure 1, with /k/ highlighted to reveal its exceptionality:

Table 5. Observed frequencies of individual consonants, and their theoretical tendencies

<table>
<thead>
<tr>
<th>#C</th>
<th>Observed frequency</th>
<th>#C</th>
<th>Observed frequency</th>
<th>#C</th>
<th>Observed frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
<td>164/++</td>
<td>/b/</td>
<td>127/+</td>
<td>/f/</td>
<td>105/++</td>
</tr>
<tr>
<td>/t/</td>
<td>134/±+</td>
<td>/d/</td>
<td>98/±+</td>
<td>/θ/</td>
<td>79/±+</td>
</tr>
<tr>
<td>/k/</td>
<td>194/−+</td>
<td>/g/</td>
<td>69/−</td>
<td>/χ/</td>
<td>60/−</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>Total</td>
<td></td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>492</td>
<td></td>
<td>294</td>
<td></td>
<td>244</td>
</tr>
</tbody>
</table>

Table 5 shows that the general pattern, indicated by +/-−, exactly matches the distribution of the other phonemes, except for /k/. Of course, the question arises...
what it is that causes /k/ to deviate from this general pattern, or, in other words, what it is that makes /k/ more successful than might be expected on the basis of the two predictions tested here. The answer has to do with a related question: why the negative effect of this positive deviation of /k/ presses harder on /g/ and /χ/ on one hand, than on /p/ and /t/ on the other; that is, /g, χ/ appear to suffer more from /k/’s success than do /p/ and /t/.

6. Discussion

/k/’s numerical superiority over /g/ and /χ/ (Table 4) is accounted for by the hypotheses of Complexity and Visibility; /g/ and /χ/ are phonemes with no fully positively marked characteristics as far as Complexity and Visibility are concerned. The triplet /k, g, χ/ is the only one of the investigated triplets that has only one member with a clear positive marker, making /k/ the most favored member of the three. That is, we already expect /k/ to surpass /g/ and /χ/. We might add that /k/ is, in spite of the lesser adroitness of the dorsum (Diver 1975: 37), not difficult to produce; only a complete closure of the velum is needed, and this is a gesture that babies already know when they learn how to swallow: lifting the dorsum. This fact may account for the exaggerated favoring of /k/ in comparison to both other stops and other dorsals.

In an absolute sense (Table 4), /p/ and /t/ are second and third best overall after /k/, so their relative frequencies can therefore be seen as the automatic arithmetic downside to /k/’s success over /g/ and /χ/, not of any weakness of their own.

One remaining question is why there is no similar favoring of /f/ in the triplet /f, θ, χ/, since also here one could say that two out of the three almost have no positively marked characteristics. The reason is that /f, θ, χ/ do not constitute a functional phonological system in Spanish, where /k, g, χ/ do, as stated above (Alarcos Llorach 1964: 52, 170–176). Therefore, the lesser preference for certain fricatives, /θ, χ/ does not necessarily have a positive effect on another one, /f/, whereas the lesser preference for the dorsals /g, χ/, does result in a greater preference for /k/, since the system in which they are organized permits the shift of one form to another, which is impossible for other triplets, like /f, θ, χ/.

7. Conclusion

The results presented in this paper may be taken, once more, as another confirmation of the hypotheses of PHB. Although the distributional facts of the phonemes in different languages are similar and hypotheses like Visibility and Number/
Complexity of articulators do function, there are also differences. These differences, however, appear to be the result of the differences in sets of phonemes and subsystems in each language, as our case of an apparent deviant distribution of /k/ in Peninsular Spanish has shown.

It is interesting to observe that in Diver (1975), where initial clusters in English are investigated, clusters with initial /k/ are more frequent than all other clusters (1975: 32). Diver does not discuss why this might be so; neither is it clear if single initial /k/ is more frequent than other single voiceless stops or other dorsals, since the data presented only show totals for all dorsals vs. apicals and labials (1975: 39).

Moreover, in Hameed (1999) it is shown that in Lucknow Urdu, dorsals score second best of all lingual articulators and almost as well as labials (1999: 3), in spite of Diver’s general statement on the lesser adroitness of the dorsum (1975: 37). Unfortunately, the individual score of /k/ cannot be deduced from the data, but this result may be taken as an indication that there are other factors that favor the production of /k/, in spite of its lesser visibility. If we consider the fact that the best scoring group, the apicals, consists of 8 members, where dorsals only have half this number, then dorsals are, relatively speaking, even the best scoring group.

Davis (1984 [1987]: 70–71) provides an argument for the relative favoring of simple /k/ over simple /t/: among the voiceless stops, /k/ is most favored by virtue of having the smallest cavity behind the constriction, therefore having the easiest time of building up pressure for an explosion (while, again, /p/ is most favored by Visibility). By contrast, among the voiced stops, /b/ is most favored (in addition to Visibility) by virtue of having the largest cavity behind the constriction, therefore having the easiest time of maintaining voicing (continuous airflow) while at the same time maintaining constriction (zero aperture). In Davis’s data on Italian, however, unclustered /k/ does not surpass unclustered /p/, as happens in our data on Spanish.

Tobin (2002) also indicates the hypothetical relevance of the size of the cavity for an unexpected favoring of /k/ vs. /p/ in certain consonant clusters in Swedish and Yiddish, since it “has the smallest oral cavity and the strongest pressure for explosion” (2002: 210). However, this hypothesis is not further elaborated, and Tobin states that “The explanation offered for these phenomena regarding the relative size of the oral cavity comes from discussions with William Diver and Joseph Davis (p.c.)” (2002: 211). It may be clear that this particular hypothesis needs further investigation.

Also the fact that language-specific characteristics may have decisive influence on the distribution of phonemes, due to historical reasons, which may be investigated by means of contrastive cross-linguistic phonological analyses, could be a factor that might contribute to the higher frequency of /k/ in our sample (Tobin 1997: 72–75).
In the particular case of Spanish, it appears that /k/ constitutes a functional triplet with /g/ and /χ/, which in itself could account for the relative favoring of the former, since the latter two are clearly more difficult to produce and therefore less favored.

It is clear that the case of /k/ deserves more study, not only in Spanish, but also in other languages. The general idea that /k/ is relatively easy to produce, which favors its occurrence as compared to other dorsals or voiceless stops for various possible reasons, is confirmed by the data of Spanish. Further study of unclustered obstruents, outside Spanish and Italian, is clearly called for.

Notes

1. The data for this paper were taken from Dekker (2000); the first author is responsible for the gathering and basic analysis of the data; the second author has re-ordered the data and is responsible for their presentation in this paper. We are indebted to the editors of this volume and two anonymous readers for helpful suggestions and corrections in my English. Remaining flaws are our own full responsibility.


3. In Peninsular Spanish, the apex may peak through the teeth while producing these sounds. Therefore, a distinction is made for this group as opposed to the other two. In /p, b, f/, visibility is undoubted, for all consonants require action of one or two lips; /k, g, χ/ are dorsals, so nothing can be seen from the outside.

References


Non-lexical interjections have been the focus of much research, but their apparent complexities, functional variations, and lack of content have led to different approaches to annotation and classification. This paper argues that they are discourse particles that function with strong cognitive linguistic bases and regularities in communication. They may have appeared as so complexly varied because they were assumed to be paralinguistic phenomena. We investigate our claims on two spontaneous speech corpora of English. In classifying these interjections into a taxonomy of discourse functions and using methods based on Phonology as Human Behavior, we find an interaction between the sound pattern of an interjection and its function in discourse, supporting our claim that non-lexical interjections are important linguistic phenomena.

1. Introduction

Our speech is replete with vocalizations like mm-hm, and so or short phrases like you know in contexts where their function or significance is disputed. Classic linguistic approaches note that these interjections do not easily fit into a single category but fall under different part-of-speech categories (e.g. adverbs, connectives, or short sentences). Work in traditional frameworks has tended to pigeonhole interjections as paralinguistic or noisy phenomena (Quirk et al. 1972) and ignore intuitions to the contrary that might indicate that these units share similar roles or functions.
An approach to language emphasizing its dynamic interactions (not rule-governed formalisms) brings us to the realization that interjections may function as signals exchanged in the process of negotiated understanding that is communication.

2. Non-lexical interjections in contemporary linguistics

Interaction strategies have often been key to understanding the contribution of paralinguistic signals (e.g. laughter and coughing) to communication. Some experts are convinced that interjections are yet another instance of a paralinguistic event (Quirk et al. 1972; Goffman 1981). Others argue for the linguistic status of interjections but have treated them primarily as aids in segmenting utterances (speech counterparts of punctuations) and in modelling the conversation flow (Hirschberg and Litman 1993; Shriberg 1994; Heeman and Byron 1998; Heeman et al. 1998; Lickley 2001).

Research in the last two decades has suggested that these units are linguistically relevant phenomena. However, linguists of different specializations and research expertise – computational (Heeman et al. 1998; Traum 1994), cognitive (Fischer 2000; Clark and Fox Tree 2002; Fox Tree and Schrock 1999), sociolinguistic (Selligman 2001), applied (Fischer 2000; Lickley 2001), and theoretical (Schiffrin 1987; Fraser 1999) – remain in disagreement about their classification both in regard to identifying what belongs in their class and in terms of finding language to describe or designate their class.

Thus, finding appropriate terminology for referring to these units has been a crux of many previous considerations. What we have called non-lexical interjections have also been classified under discourse markers, discourse particles, and cues. However, for example, no consensus exists on what items or phenomena should be included in the class discourse particles. Nonetheless, even advocates of differing terminologies agree that they are studying linguistic phenomena: language devices which are generally syntactically detached and help elucidate the discourse structure.

With further complications, non-lexical interjections seem lexicalized in the sense that they have accepted denotations. The non-lexical status of these items may result from the variation they seem licensed to undergo; a license not as freely given to more commonly accepted lexical items. Mm-hm, mmm-hmmmm, mmmmm, and uh-huh, all seem to function much like (the indisputably lexical) yes. Whether or not all of these variants are really the same, these categorization difficulties and the lack of a predictable structure for interjections contribute to arguments for their non-lexical and paralinguistic status.
Chapter 7. Sound patterns of English non-lexical interjections

We will argue that non-lexical interjections are more regular than previously thought: their variants can be simplified to a base form. Any study of interjections must take into account how they can function as a part of discourse. Moreover, consistently ignoring the discourse function of these terms itself creates and exaggerates the impression that their sound makeup is unstructured.

Our analysis supporting the functional character of non-lexical interjections will draw on principles from Phonology as Human Behavior (PHB) (Diver 1979; Davis 1984 [1987]; Diver 1995; Tobin 1997). Our recognition of the importance of communication and human perceptual and production constraints on language accords well with the “human factor” stressed in PHB and Columbia School Linguistics generally.

3. The phonology of non-lexical interjections

Little research has been done on the phonology of non-lexical interjections. Most of the research has been informal observations or speculative commentary. The more focused studies have been based, predominantly, on sound symbolism. For example, researchers have casually but consistently noted that non-lexical interjections commonly involve infrequent or illegal phonotactic combinations (Abelin 1999; Montes 1999; Ward 2000). They have also noticed that interjections in different languages share phonetic similarities.

In examining Icelandic, English, Polish, Hungarian, Finnish, Ososo, Malagasi, and Slovenian interjections, Abelin (1999) found that non-lexical interjections involve mostly labial or alveolar sounds. She also argued that their sounds have intrinsic (iconic) meanings. Like Abelin, Ward (2000) appeals to sound symbolism. In studies of American English (Ward 2000) and Japanese (Ward 1998), he proposes that the sound structure of non-lexical interjections (or “grunts”, as he terms them) reflects the attitude or knowledge of the speaker. In a more detailed phonetic analysis of American English interjections, Ward (2006) builds further on his argument that the meanings of grunts may be predicted given their phonetic characteristics. He speculates that these phonetic characteristics perhaps emerge from cognitive factors. However, he proposes very specific sound symbolic mappings and does not focus on the motivations of the relationships among form, function, and meaning.

Rather than submitting to sound symbolism, we argue that the phonotactic skewings have clear cognitive bases. In the style of PHB, we can identify functional motivations for the sound structure of non-lexical interjections in terms of the “grounding process” (a term coined by Clark and Schaefer, 1989), where “grounding” is the negotiation of understanding among conversation participants. In so doing, we will gain insight into the role of non-lexical interjections in discourse.
4. Interjection sound patterns: The hypotheses

PHB posits that the non-random sounds in languages emerge from the (sometimes) conflicting goals of minimizing production efforts while negotiating efficient communication (cf. Tobin, this volume). We can reason then that less crucial changes in the discourse interaction will involve non-lexical interjections with less marked sounds and sound patterns. For example, we can explain why most pause fillers (the type of non-lexical interjections used to keep other participants from interrupting) seem to consist primarily of vowels (e.g. uh, ah, eh, oh). Vowels require less effort to articulate than consonants. The pause fillers are intermediary moments between heavier, information-laden articulatory gestures; it seems natural for the speaker to expend only minimal effort in their production, just enough to hold the floor (or not).

A relative scale of articulatory difficulty and perceptual saliency for particular sound patterns would, we think, indicate the degree to which a speaker desires to hold the floor. PHB and biomechanical studies (see Kim 1995 for a brief review) agree that the apex of the tongue and the lips are among the most flexible oral articulators. This explains Abelin’s (1999) observation that interjections generally tend to involve sounds produced by either the lips or the apex of the tongue.

More formally, we can consider the discourse functions of non-lexical interjections to be either static or dynamic. Those that are static do not change the current beliefs or understanding of the participants or the intentional direction of the topic flow, but indicate, for example, the speaker’s attendance to the conversation. We hypothesize that these static-functioning non-lexical interjections will overall be much simpler and vary less phonetically than interjections indicating more dynamic participation, which unlike static interactions, updates the beliefs and knowledge of the participants. In other words, static-function interjections will tend to involve the less salient sounds and/or more easily articulated sounds (what will be referred to as unmarked sounds). This implies a comparatively limited phonetic inventory and very simple syllable structures (most likely monosyllables). Analogously, we hypothesize that dynamic-functioning interjections indicate a speaker’s willingness to increase articulatory effort for greater communicative holds and/or to produce particles with greater perceptual distinctions (or marked sounds).

5. DiSPEL (DIScourse Particle Expert model): A functional taxonomy

In order to test these hypotheses, we created a functional taxonomy for interjections that is simple enough for computational purposes but which also sufficiently captures the functions of interjections as discourse particles.
As discourse particles, non-lexical interjections constrain the discourse process. They represent the speaking intentions of participants, making utterances more predictable and thus understandable. A trend in discourse particle models has been to separate the discourse-managing processes (the process of grounding) from the flow of topic information. Within our DiSPEL framework all discourse particles, including non-lexical interjections, function as grounding units. Our taxonomy, developed from one presented by Traum (1994), also treats the grounding process as the set of possible contributions constrained by the information and intention of the participants.

Four factors in DiSPEL describe the management of communication or discourse moves, as shown in Figure 1.

<table>
<thead>
<tr>
<th>Discourse Moves</th>
<th>Topic</th>
<th>Relation</th>
<th>Projection</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgement</td>
<td>old</td>
<td>initiating</td>
<td>other</td>
<td>negative</td>
</tr>
<tr>
<td>Expansion</td>
<td>new</td>
<td>related</td>
<td>self, other</td>
<td>neutral</td>
</tr>
<tr>
<td>Correction</td>
<td>old</td>
<td>initiating</td>
<td>self, other</td>
<td>negative</td>
</tr>
</tbody>
</table>

**Figure 1.** DiSPEL discourse management moves

1. **Topic** indicates whether the utterance is related to a “new” or a previously spoken-of (“old”) topic.
2. **Relation** is the hierarchical interdependency among the utterances in a dialogue. Dialogues include main topics and subtopics each involving dependencies among the utterances. Utterances relate to each other not only by topic but also as the result of the communication strategies speakers use. For example, questions expect answers; a greeting is typically returned. Answers and return greetings are “related” to previous utterances. After greetings are exchanged, one of the participants would often “initiate” a new topic.
3. **Projection** indicates whether a speaker is making a statement based on his own knowledge (“self”) or based on something the “other” participant has said.
4. **Evaluation** is the type of assessment a speaker gives: that is, “positive” (affirmation), “negative” (disagreement), or “neutral” (registering attention).

These four factors delimit the moves that discourse particles realize. According to grounding theory, the moves monitor and manage the discourse process. We have...
classed the examples in our study based on three discourse moves: (1) **acknowledgment**, (2) **expansion**, and (3) **correction** (negative assessment).

**Acknowledgment** indicates the assessment of information provided by the other speaker (thus projection is always “other”). The type of evaluation an acknowledgment move has (positive, negative or neutral) dictates the types of utterances that are expected to follow. For example, *uh-huh* (from the TRAINS 91 corpus, Allen & Schubert 1991), constitutes a positive acknowledgment move as it indicates that speaker M agrees with speaker S’s utterance. Cases in which *uh-huh* is followed by a negation like *but* are also classified as acknowledgement (even if partial) because the negation indicates a change in turn.

(1) S: that only takes a couple of minutes  
M: *uh-huh*

Negative acknowledgment describes disagreement. *Uhh* in this TRAINS 91 example indicates dissent:

(2) M: if there are oranges at the orange juice factory  
S: *uhh* no

Neutral acknowledgment describes attentiveness (assessment without agreement or disagreement). In Example (3), also from TRAINS 91, speaker S is suggesting a plan to participant M, who utters *m-hm* to encourage speaker S to continue rather than to actively agree or disagree with the plan.

(3) S: with this plan we can get the bananas to Corning at 1  
M: *m-hm*  
S: and we have the OJ in Corning at 6  
M: okay

**Expansion** moves (by a speaker or an interlocutor) extend a previously spoken topic with new information. An example is provided in (4):

(4) M: *so um* do I tell you what to do at this point

Lastly, **Correction** is a move that partially or entirely changes an utterance or topic that has been discussed. It is always a negative type of evaluation. In the following example, both *oh* and *whoops* function as self-corrections:

(5) M: keep the engine there if it’s not needed anywhere else send  
S: well we only have oranges in Corning  
M: *oh* I’m sorry right *whoops*

*Oh* and *whoops* signal speaker M’s realization of having overlooked part of the situation when giving an initial plan.
For our tests, we classed interjections first according to the discourse move taxonomy and then distinguished between those that function statically or dynamically in the discourse. We tested our hypotheses in two dialects of English.

6. The corpora: Choosing and post-coding

Finding appropriately annotated public-domain corpora with consistently annotated non-lexical interjections was difficult because early research in spoken language systems filtered out interjections as irrelevant to the process of communication. This problem restricted our corpus choice to the TRAINS 91 dialogues (Allen and Schubert 1991) and the HCRC Map Task Corpus (Boyle 1990, Carletta et al. 1996).

The TRAINS 91 corpus is a collection of 16 task-oriented Wizard of Oz game dialogues among gender-balanced and regional-dialect-balanced American English speakers. One speaker assumes the role of a “task manager” who must interact with the other speaker (the Wizard of Oz) who alone knows the conditions and rules of the task to be completed. Dialogues ranged from 40 seconds to 13 minutes in length.

Although not phonetically transcribed, TRAINS 91 includes orthographic transcriptions of the variations (e.g., ohhh) of base forms (e.g., oh) which approximate the token articulation of the given interjection. The transcription also includes overlapping speech. We semi-automatically tagged the non-lexical interjections and annotated them with the grounding moves of the DiSPEL taxonomy. We also noted the interjections’ locations: that is, where in a turn they occurred (a turn being a speaker’s contribution, which may include more than one utterance).

The HCRC Map Task corpus consists of collaborative problem-solving task dialogues between pairs of speakers. Each speaker had maps with slightly different landmarks (neither participant was able to see the other’s map). One of the pair would instruct the other towards a particular destination (Boyle 1990). The HCRC Map Task corpus consists of 128 digitally-recorded dialogues spoken by a gender-balanced pool of 64 Glaswegian English speakers.

Although the orthographic transcription of Map Task did not reflect the token-type articulation of each non-lexical interjection, its transcriptions were time-aligned to audio files. With these, we did a broad phonetic transcription of the non-lexical interjections which occurred. The corpus also provided various codings including dialogue move and part-of-speech annotations that facilitated automated processing of data. The dialogue moves used in the Map Task annotations were not entirely grounding moves; thus, it would have been inappropriate to use them directly in our study. We compensated by automatically mapping them to the DiSPEL moves by also taking into account their part-of-speech annotations.
7. Markedness and unmarkedness

Our hypotheses lead to another sub-problem: we need to identify and define unmarked and marked sounds. Like discourse particles, the concept and definitions of markedness have a long reputation of controversy. Here, we are sympathetic to Greenberg’s reinterpretation of this Prague School of Linguistics concept (Greenberg 1966, 1975).

In the Prague School of Linguistics, markedness referred to features that contribute to phonemic oppositions; it was what distinguished one phoneme from another. Hence markedness can be language-specific. Greenberg explicitly argued for a more functional or usage-based approach: markedness can have motivated cross-linguistic similarities. Items that are the most frequent, the simplest, the most universal, acquired the earliest by children, developed earliest in the language and contrast least with other unmarked categories are often considered unmarked. For us, this characterization implies that markedness can have cross-linguistic similarities as well as be language and even dialect-specific. Some perceptual and production advantages are fundamental to physiological capabilities (linguistically universal), whereas others may be conditioned to the contrasts present (the needs of) specific languages and dialects. Speakers of one dialect, for example, can maintain categorization (perceptual) boundaries and phonetic variations in articulation that differ from those maintained by speakers of another dialect. As such differences can exist from dialect to dialect, we had to identify the marked sounds relative to both American English and Glaswegian English.

Specifically for our task, we define markedness as characterized by the increased effort required for production and/or by its relatively greater perceptual salience. We looked at markedness only at the phone and syllable levels. We defined as unmarked any sounds that:

- Occur in neutral or rest positions before or between articulations. This is predominantly a physiologically-based criterion and should show cross-linguistic similarity.
- Are typical sounds occurring in unstressed syllables. Unstressed syllables, as explained with principles of PHB and as implied in psycholinguistic experiments (for example, Jusczyk et al. 1993), carry a low communication load. Since they convey relatively less information than stressed syllables, listeners do not need to attend as much to them. Less attention paid to unstressed syllables also means that less effort is needed. On this basis, unstressed syllables should involve sounds which are not as distinctive perceptually or production-wise. As phonological systems change from language to language, this criterion uncovers primarily language-specific and even dialect-specific differences.
Chapter 7. Sound patterns of English non-lexical interjections

- *Have low syllable complexity.* Syllable complexity can involve both cross-linguistic similarities and language-/dialect-specific differences. However, we only looked at cross-linguistic similarities in measuring syllable complexity by the presence of consonant clusters and the number of syllables. For example monosyllables, compared to multi-syllabic items, require less effort and are less salient perceptually. They are thus unmarked.

These criteria are not meant to be exhaustive. For example, non-average durations and pitch are also likely to indicate marked sounds, but analyses of these prosodic qualities and markedness on super-syllable levels remain to be investigated.

Given the criteria above, we reviewed the sounds that would be marked/unmarked for American and Glaswegian English. The lips when stably closed seem physiologically the most neutral non-articulating static position between utterances. That is, the resting articulatory position seems to correspond to /m/ without phonation and appears to be linguistically universal (Klaus Kohler 2003, personal communication). We therefore classed /m/ as unmarked for both American and Glaswegian English. As markedness is not always linguistically universal, we also classified markedness of sounds relative to each dialect.

The schwa is the most central position for an American English speaker, as well as a very typical vowel in unstressed and reduced syllables. Thus, for our analyses, we classed /m/ and /ə/ as the unmarked sounds of American English. Marked sounds then include rounded vowels (e.g. /o/), lengthened vocalizations (long vowels or mmmmm), non-central or tense vowels, and non-sonorants (such as stops).

For Glaswegian English, we looked to Abercrombie (1979) for the typical vowels it had in unstressed syllables. Figure 2 presents the typical vowels for Standard Scottish English (SSE) and the counterpart pronunciation in standard British English (Received Pronunciation, RP).

![Table showing vowel pronunciation in SSE and RP](image)

**Figure 2.** SSE Vowels typical in unstressed syllables (Abercrombie 1979)

For both dialects, multi-syllables and consonant clusters were considered marked.
8. Results

The ten dialogues of TRAINS 91 included 410 tokens of non-lexical interjections. We analyzed six of the eight dialogues in the HCRC Map Task corpus, thus processing 3175 tokens of non-lexical interjections. For the TRAINS corpus, we first identified base forms by their relative high frequencies in contrastive functional realizations (as based on the taxonomy in Figure 1) and their locations in turns. In correlating the variations of the interjections to their base forms, articulatory or sound similarities are insufficient criteria because interjections with different functional realizations often have close sound structures.

We found that the patterns of frequencies in functional realizations, in addition to sound proximity, provided a reliable method of identifying variants of base forms even though frequencies were very low for some forms. As can be seen in Figure 3, in some cases we have left very low-frequency items (in parentheses) as base forms because of the possibility that they have functions that may be distinctive enough to warrant their status as base forms. The intonation and other prosodic features will most likely be key to aiding this distinction. These issues will require further investigation and are out of the scope of this chapter.

<table>
<thead>
<tr>
<th>Base-Forms</th>
<th>Variants</th>
<th>Moves</th>
<th>Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ah</td>
<td>(hmm)</td>
<td></td>
<td>2,1</td>
</tr>
<tr>
<td>(aha)</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>(eh)</td>
<td></td>
<td></td>
<td>1 / 2</td>
</tr>
<tr>
<td>(err)</td>
<td></td>
<td></td>
<td>2, 1</td>
</tr>
<tr>
<td>m-hm</td>
<td></td>
<td></td>
<td>0, 1</td>
</tr>
<tr>
<td>uh</td>
<td>uhm, uhh, (uhhm), (uhhh), (uhmm)</td>
<td>Exp, Ack, Corr</td>
<td>2, 1, 3, 0</td>
</tr>
<tr>
<td>um</td>
<td>uhm, umm, (uumm), (uhhm)</td>
<td>Exp, Ack, Corr</td>
<td>2, 1, 3, 0</td>
</tr>
<tr>
<td>uh-huh</td>
<td></td>
<td></td>
<td>0, 2</td>
</tr>
<tr>
<td>hm</td>
<td>(hmm), (m), (mm)</td>
<td>Corr</td>
<td>2, 1</td>
</tr>
<tr>
<td>oh</td>
<td>(o), (ohh), (oo), (ooh), (oooh)</td>
<td>Corr, Ack, Exp</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>(oops)</td>
<td>(whoops)</td>
<td>Corr</td>
<td>0, 1 / 2</td>
</tr>
<tr>
<td>(ouch)</td>
<td>(uch)</td>
<td>Corr</td>
<td>0</td>
</tr>
<tr>
<td>wow</td>
<td></td>
<td>Corr</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 3. TRAINS corpus: non-lexical interjections and their variants (see text for explanation)
In Figure 3, interjections and their variants are listed in decreasing order of frequency. Items in parentheses indicate very low total frequency (across base forms if a variant). Items in italics are sound synonyms, that is, variants that occur with different baseforms. Location indicates turn information: 0=constitutes turn, 1=at the beginning of turn, 2=within turn, 3=at the end of the turn.

Our data provide some support that interjections are context-dependent and that their function depends on a combination of their position in turns, their denotation, and the context in which they appear.

Looking at the functional distribution of the non-lexical interjections, the first observation is that their most frequent position is within the body of the turn (“2”). However, most of these were functionally self-expansion interjections (Exp), which indicate the current speaker’s intention to further expand the utterance by contributing more information. The least frequent position is at the end of the turn (“3”). Therefore, in general, interjections appear to prepare the listener in predicting the following utterances. The 2% that occur at the end are predominantly interjections which speakers use for self-expansion (indicating an intended beginning of a turn) but were interrupted by the listener.

The second most frequent function of interjections is to indicate change or correction. The change could be topical, indicating an initiation of a new topic. Or it could be an indication of a speaker’s knowledge update. These are all tagged as self-repair or self-realization (Corr, self). Continuing the general trend of location distributions, this type of interjection tended to appear in the body of the turn (“2”). The least frequent function is that of acknowledgment (Ack).

In support of our hypotheses, results show that the degree of markedness in the sound makeup of interjections relates directly to the degree of interaction (see Figure 4); this is as we hypothesized. The data also confirm that the syllable structure of static-functioning interjections tends away from multi-syllabic forms (0.5% multi-syllables) more than the dynamic-functioning ones (3.4% multi-syllables).
The only sound which appeared not to match our predictions is the lengthened /m/; we had assumed that it is marked, yet it functions statically more frequently. However, a difference does exist between lengthened /m/ and its shorter form. The lengthened /m/ occurs primarily at the beginning and the end of turns (thus marking the change in turns) whereas the shorter form occurs primarily within turns. This may imply that the sound structure of non-lexical interjections depends on both function and location and supports our hypothesis that marked sounds indicate the dynamics of interactions.

9. Interjections in the HCRC Map Task corpus

Even a first pass of the Map Task analysis (in Figure 5) attests to our supposition that sound markedness is significantly interrelated with discourse dynamics.
The results for the Map Task corpus are slightly more complicated than for TRAINS as there were many more multi-syllabic forms in the Map Task corpus (probably due to the nature of the task). Whereas the TRAINS analyses involve markedness that was primarily at the phone level, the Map Task results also reflect markedness on the super-phone level.

To examine the effect of syllable complexity as a criterion of markedness, we split the data in Figure 6 to show that taking into account the number of syllables also provides support for our hypothesis that static-functioning interjections too have unmarked sound structures and further strengthens our claims for sound-function interdependency.

Splitting the data according to syllabic complexity reveals that all multi-syllabic non-lexical interjections in Map Task have dynamic functions. This may indicate a stronger support for our hypotheses; however, this may also only mean that the HCRC Map Task corpus did not have multi-syllabic non-lexical interjections which functioned statically because of the nature of the task used during data collection. Nonetheless, the role of number of syllables becomes more convincing with a detailed look at a subset of the data.

Figure 5. Map Task: non-lexical interjection sound structure vs. discourse dynamics
Figure 6. Map Task: syllable complexity in monosyllabic interjections (top) and in multisyllabic interjections (bottom)
Figure 7 presents the data for only the unmarked sounds in Standard Scottish English. Contrary to our hypotheses, the unmarked sounds /m/ and /ʌ/ seem to occur in dynamic-functioning items instead.

![Figure 7](map_task_markedness_by_discourse_dynamics_for_unmarkedPhones.png)

**Figure 7.** Map Task: markedness by discourse dynamics for unmarked phones

However, when we take into account syllable complexity, the data fit our hypotheses again. All the instances in which the unmarked /m/ had the “unexpected” occurrence in dynamic functions were because they occurred in bisyllabic items. Over half of the deviations involving the unmarked /ʌ/ can also be explained in the same way (see Figure 8): it is mostly dynamic when multisyllabic and less so when monosyllabic.

Analyzing the remaining deviant /ʌ/-dynamic function occurrences, we found that almost all of these instances were characterized by extremely short durations (average duration of 85 msec with a standard deviation of 26) and orthographically transcribed as *oh*. This might be that the target was /o/ but the sequence was too short for the acoustics to be perceived as the intended articulation, and it also indicates that duration analysis, as well as other prosodic investigations, may be insightful.
Figure 8. Map Task: effects of syllable complexity for unmarked phones: monosyllabic (top) and multi-syllabic (bottom)
Chapter 7. Sound patterns of English non-lexical interjections

10. Conclusions

The goal of our analyses was multi-fold. Our main pursuit was to find support for our claims of the importance of non-lexical interjections in linguistic analyses and to uncover their regularities. Following the corpus linguistic method of Phonology as Human Behavior, we tagged our non-lexical interjection data into factors that reflect the importance of the role of communication (in our case, as reflected in the discourse structure) and human constraints in communication (sound markedness). With these classifications we are able to find sound structure in non-lexical interjections which does not differ from the phonological principles of the dialect in which they occur. That is, like other sounds in a language, non-lexical interjections reflect the balance of human constraints in communication. Perhaps non-lexical interjections reflect this balancing act more transparently than other more lexicalized items.

Similar studies across different languages and their dialects would certainly be necessary, but the results are promising and reinforce the cognitive linguistic tenets that

– the apparent “noise” in language can actually be one of the most interesting and insightful windows into language phenomena;
– meaning, form and function are inseparable; and
– structure emerges when considering the factors and context impinging on language use.

Acknowledgements

We would like to thank our editors and reviewers for their invaluable comments. The ideas which Yishai Tobin and Liz Shriberg shared with us were important, but above all, this work could not have developed further without their encouragement. Thanks also to Matthew Aylett for help with the HCRC Map Task corpus.

References


CHAPTER 8

Phonology without the phoneme

Joseph Davis
The City College of New York

In his otherwise radically innovative linguistics (Columbia School), William Diver retained the classical phoneme, defined on the basis of contrastive distribution. He did so despite his rejection of most of the apparatus of traditional, descriptivist, and contemporary linguistics, and despite well-known analytical difficulties. Diver evidently saw the phoneme as being required on theoretical grounds, specifically the communicative orientation. Communication, however, does not require contrastive segmental units, and Columbia School phonology need not rely upon the phoneme, which is superfluous to its findings anyway.

1. The phoneme in Diver and Columbia School

The phoneme occupies an unusual position within Columbia School linguistics; it may be one of only two linguistic constructs to survive essentially unaltered in what is otherwise a radically innovative theory. Diver (1980, 1981, 1995), like Saussure (1916 [1972]) before him, excluded from his linguistics the traditional sentence and all its parts. Though Diver did adopt a version of the linguistic sign, Diver’s linguistic signal differs in essential ways both from Saussure’s signe and from earlier versions of this fundamental semiotic unit (Davis 2004; cf. Reid, this volume). Practically none of the linguistic machinery devised in Europe or America between the times of Saussure and Diver survives in Diver’s theory: the word class, the sememe, the binary feature, markedness, the archiphoneme, the allophone, and so forth, all fail to appear in Diver. And, of course, Diver virtually ignored all the innovations of theories contemporary with his, including generative grammar and discourse and cognitive linguistics as well.

Two inherited devices do survive in Diver: “The successful techniques of descriptive linguistics... provide us with the identity of a certain list of distinctive
units, phonemes and morphemes” (Diver 1995: 54). We pass over, for now, the question of whether those techniques were in fact successful. For Diver, phonemes and morphemes were not ends in a descriptive linguistics but means to an explanatory linguistics.

[The descriptivist techniques] do not... provide us with a statement of the characteristics of these units, and, hence, they do not provide the kind of explanation that we are looking for....

...

... What needs to be added to these bare initial hypotheses is a statement of how the distinctive unit of sound, the phoneme, is to be related to the sounds themselves, and how the distinctive unit of form, the morpheme, is to be related to the communication. (Diver 1995: 54–55)

For Diver (1995: 55), “The constant problem is, why these units in this position and why this order? What is the motivation?” Linguistics for Diver always requires explaining distributions. The account of the distribution of phonemes is phonology; that of the distribution of morphemes is grammar. Diver explained the distribution of phonemes by assigning to them appropriate phonetic substance; he called these explanatory constructs *phonological units*. He explained the distribution of morphemes by assigning to them appropriate semantic substance; he called these explanatory constructs *signals*. The present paper will not undertake to question the retention of the morpheme in Diver’s linguistics; it will scrutinize only the appropriation by Diver of the descriptivist phoneme.

2. The phoneme defined

Informally defined, the phoneme is a distinctive unit of sound. One influential and early attempt at formal definition is that of Jones (1932 [1966]: 31): “A family of sounds in a given language, which are related in character and are such that no one of them ever occurs in the same surroundings as any other in words.” This phonemic principle can easily be illustrated with the words *keep, cop,* and *coop.* The initial sound of *keep* is formed in a relatively forward position in the mouth; the initial of *cop* is formed further back; and the initial of *coop* is also back but has the lips rounded. So the three sounds are different, but they are “related in character” in that they are all aspirated voiceless stops made with the back of the tongue somewhere near the velum. Furthermore, the three cannot occur in each other’s “surroundings”: the palatal [k] of *keep* never appears in English directly before the [a] sound of *cop,* and so forth. The three k’s, then, illustrate phonetic similarity and
complementary distribution, the two pillars of phonemic status. The three make one phoneme, /k/.

If either of these two pillars is absent, then phonemic status does not hold for a group of sounds. For example, if two similar sounds can appear in the same environment – that is, if they are not in complementary distribution – they are distinctive, and so they are different phonemes, as the /k/ and /g/ in cool and ghoul. On the other hand, two sounds which are in complementary distribution are not considered one phoneme if they are phonetically too dissimilar, the classic cases being exemplified by English hang. [h] in English can occur only prevocally and [ŋ] can occur only postvocally, but they are not grouped as one phoneme because they are deemed phonetically too dissimilar, sharing nothing but consonantal status. Of course, that begs the question, How similar do segments have to be in order to be classified as allophones of a single phoneme? The phonemic procedure begins to look less than algorithmic.

3. Analytical difficulties with the phoneme

Analytical difficulties with the phoneme arise immediately and resist satisfactory solution. Nor are they unusual, popping up in language after language. Two illustrations will suffice.

Consider two competing analyses of the English series beat, bit, bait, bet (Table 1). Do these words determine two or four phonemes?

<table>
<thead>
<tr>
<th>Table 1. Competing phonemicizations in English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>beat</td>
</tr>
<tr>
<td>bit</td>
</tr>
<tr>
<td>bait</td>
</tr>
<tr>
<td>bet</td>
</tr>
</tbody>
</table>

Assuming that the status of /y/ is independently justified (yes /yes/, S /ês/; kite /kayt/, cot /kat/), the two analytical possibilities are for four phonemes (/i, ɪ, e, ɛ/) or for two (/i, ɛ/). Both analyses successfully describe the distinctions maintained by the language. The four-phoneme analysis yields simpler, i.e. shorter, transcriptions. The two-phoneme analysis yields a simpler, i.e. smaller, phonemic inventory. Some sort of simplicity metric is required.

Chao (1934) presented the definitive phonemic conundrum, titling his paper, in fact, “The Non-Uniqueness of Phonemic Solutions of Phonetic Systems.” Man-
darin has four fricatives, here labeled abstractly \([x_1, x_2, x_3, x_4]\), which exhibit the following distribution:

- \([x_1]\) occurs before \([i], [ü]\)
- \([x_2]\) occurs elsewhere (not before \([i], [ü]\))
- \([x_3]\) occurs elsewhere (not before \([i], [ü]\))
- \([x_4]\) occurs elsewhere (not before \([i], [ü]\))

\([x_1]\) appears only before high palatal vowels; the three others, \([x_2, x_3, x_4]\), appear before all other vowels. Clearly, \([x_2, x_3, x_4]\) are distinctive and must be separate phonemes. But \([x_1]\), being phonetically similar and in complementary distribution with the other three, must be assigned to one of them, forming a two-member “family” with it. But with which one? Chao’s point was that it is impossible to say, and that, moreover, this state of affairs is commonplace: “given the sounds of a language, there are usually more than one possible way of reducing them to a system of phonemes.”

Above and beyond the problems with such segmental phonemes, there was always the question of what to do with distinctive suprasegmentals. Should the distinction between Italian *casa* ‘house’ and *cassa* ‘box’ be characterized as involving repeated segments (/kaasa, kassa/), or a non-segmental phoneme of length? How to handle the Mandarin quadruplet /ma/ ‘mother, hemp, horse, scold’? With paradigmatic sets of phonemes (/a\(^1\), a\(^2\), a\(^3\), a\(^4\)/), or with non-segmental phonemes of pitch that more accurately describe the pitch pattern across the entire stretch of voicing in the word (i.e., including the [m])? What to do with the English pair *abstract*? Have a pseudo-segmental phoneme of stress (/’æbstrækt, æb’strækt/), or a non-segmental phoneme of stress? Clearly, segmental phonemes, one coming after another, cannot be the whole picture.

Twaddell (1935) argued that the phoneme could not reliably be said to have either psychological or physical reality and so could only be “an abstractional, fictitious unit.” These sorts of difficulties kept piling up until Hockett (1947) made use of simultaneous features rather than sequential segments, an approach that Joos (1958 [1963]) characterized as “phonemics without the phoneme.” The theoretical status of the phoneme has ever since been in doubt.

4. Diver’s theoretical justification for the phoneme

Though aware of the difficulties with phonemicization, Diver retained the phoneme on theoretical grounds. The phoneme, he evidently believed, is required by the communicative orientation in this theory, by the need for distinct signals. And
so, in his seminal phonological work, Diver (1979) was led to minimize the analytical problems in the procedure.

There has never been any great difficulty, either for modern phonemicists or ancient alphabet makers, in establishing the existence and identity of phonemes. The area of disagreement in phonemic analysis is small, and the task of reducing languages to writing does not produce wildly disparate results. (Diver 1979: 162)

Diver here is not being casual about the phoneme; it is not merely a makeshift on the way to his explanatory phonological unit. Here the phoneme is a linguistic unit according to the essential Diverian criterion: it is the solution to a problem involving observables.

[A] skewing in the phonetics pose[s] a problem that, against the background of the communicative factor, has led to the hypothesizing of phonemes, the phoneme thus being the solution to the problem posed by the phonetic skewing. (Diver 1979: 167, emphasis added, JD)

Diver (1979: 169) even dares to suggest psychological reality for the phoneme: “Users of a language behave as though they have learned certain distinctive units, the phonemes, which they deploy for communicative purposes.” It is fairly clear, then, that Diver (1979) accepts the classical phoneme as a theoretical unit.

Diver’s contribution to the field at this point is to go beyond the phonemic problem to solve the problem posed by the non-uniform distribution of phonemes within signals:

Given a stock of distinctive units and a large corpus of phonemic transcription, it might be expected that, other things being equal, the phonemes would be distributed symmetrically within the larger linguistic units that they serve to differentiate.... We know, of course, that they are not distributed symmetrically, and it is the motivation for the particular skewing we find that constitutes the phonological problem. (Diver 1979: 167–168, emphasis added, JD)

It is the phonological problem – explaining the distribution of phonemes – to which Diver devotes most of his attention (and for which he deserves most credit). Fittingly, then, he is remembered not so much for handling phonemes but for discovering the characteristics of what he termed phonological units, which are phonemes clothed with explanatory phonetic and human-factor characteristics. For example, analysis of the distribution of /p, t, k, b, d, g/ compels him to characterize them in terms of their active articulators rather than their passive place of articulation. But it in no way minimizes these achievements to note that, along the way, Diver does fully retain the classical phoneme.
Even at his most radical, in his final paper, Diver (1995) has a phoneme.  

The hypotheses closest to the observations are those normally provided by a phonemic analysis (or by the learning of the use of an alphabet); that is, an inventory of distinct sounds used differentially for the purpose of communication. (Diver 1995: 111, emphasis added, JD)

Note that Diver (1995) appears to waive the criterion of phonetic similarity; “distinct sounds used differentially” would presumably be those established by minimal pairs such as cool and ghoul, that is, those established distributionally. It is not clear what Diver (1995) would do with hang.

A careful reader will be troubled by the survival of this inherited construct in such a radical linguistics as Diver’s. Even more troubling is Diver’s uncharacteristic nonchalance with regard to those pesky phonemics problems. This is the same Diver (1995) who writes, “To clear the air, then, the term ‘theory’ will here be used to refer to a summary of the general characteristics of successful solutions to individual problems” (p. 45, emphasis added, JD); and who writes, “To avoid metaphysics, then, one simple technique is purely analytical: abandon hypotheses that don’t work” (p. 47).

The phoneme is a hypothesis that doesn’t work. Or perhaps it works only about as well as other traditional and modern linguistic notions: sentence, word class, and so forth. It does not meet Diverian standards for analytical success.

If the phoneme fails as a hypothesis, and if Diver knew it had not been entirely successful, why did he keep it in his arsenal? Diver apparently believed that the communicative orientation, so well established in his theory, necessitated the existence of the phoneme: the phoneme is what makes distinct signals possible. And so whatever difficulties are encountered in the identification of phonemes must be mere impediments testing our analytical ingenuity; they do not disconfirm the phoneme as a theoretical construct.

5. The theoretical justification questioned

In fact, however, the communicative orientation does not require the phoneme. Communication merely requires distinct signals, and distinct signals can be achieved without distinct segments that are defined by complementary distribution and phonetic (i.e. substantive) similarity. For example, in a communication system, signals may be entirely distinct, without segments. Such is the traffic light, where the signal for ‘stop’ is a thoroughly different color than the signal for ‘go,’ its duration not segmented into a sequence of colors.
Chapter 8. Phonology without the phoneme

The logical alternative to signals being entirely distinct – distinct throughout their duration – is for them to be *partly* distinct – distinct in only parts, or intervals, of their duration. This appears to be what language does. Still, even if a communication system does have intervals, or segments, these need not be identifiable through complementary distribution and substantive similarity.

Consider the hypothetical nonlinguistic system of communication shown in Figure 1. It has a lexicon of ten (1–10) distinct graphic lexemes (the meanings of the lexemes are immaterial here).

**Lexicon of 10 lexemes:**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>$ §O</td>
<td>—O</td>
<td>—</td>
<td>$</td>
<td>—</td>
<td>—</td>
<td>§-O§</td>
<td>D§-</td>
<td>—D</td>
</tr>
</tbody>
</table>

*Figure 1.* A hypothetical graphic communication system

By applying an emic procedure, we can easily identify four graphemes: vertical /|/, seahorse /§/, horizontal /-/ (with the allograph [ — ] in the environment of vertical /|/), and circle /O/. For example, the minimal pair of lexemes 1 and 2 allows us to identify the graphemes horizontal /-/ and circle /O/; these distinguish lexemes 1 and 2. We can also identify allographs. Horizontal has two forms: a long variant [ — ] found in the immediate environment of /|/ and a short form [-] found elsewhere (except lexeme 10). So far, so good for the descriptivist procedure. But what are we to do with the stirrup iron [D] that appears in lexemes 9 and 10? Comparing lexemes 1 and 9 would give stirrup iron [D] in initial position as distinct from vertical /|/ in initial position. So [D] might be a grapheme unto itself, /D/. Alternatively, [D] might be analyzed as a fused sequence of /|/ and /O/; notice that /|/ and /O/ never co-occur. This analysis makes sense, too, in terms of the visual substances involved. Stirrup iron [D] sort of looks like a combination of vertical /|/ and circle /O/, in that order. (And /O/ never occurs initially, further confirmation of the order.). Note too that horizontal takes its long form [ — ] in the immediate environment of [D], suggesting that [D] is really /|/ plus /O/. This, then, is a perfectly workable system of communication whose intra-signal units cannot be identified on purely distributional and visual grounds.

Similarly with a real written system of communication, the Greek alphabet. The letter written ς appears only before a space, while the letters written σ and ζ, among others, never appear before a space.

σκληρόν σοι πρός κέντρα λακτίζειν
If we wanted to assign ζ as a variant, or allograph, of one of the other letters, we would not be able to do so purely on the basis of distribution. Or, in the formulation of Diver (1995), are ζ and ξ distinct graphs used differentially for the purpose of communication? Not any more so than ζ and σ. The Greek lexicon contains no minimal pairs involving either contrast.

We are thus led to recognize that the communicative orientation does not require the existence of phonemes. If that is the case, then the phoneme must earn its way into the theory, and the only way it can do that in accordance with Columbia School principles is through analytical success. Yet, as we saw in our historical tour through the descriptivist procedure, there has not been analytical success for the phoneme.

One phonologist who takes seriously the failure of phonemics is Bybee (2001). Her work is particularly interesting here because it is bolstered by evidence and argument that Columbia School linguists would recognize as an appeal to a human factor, phonetics, and communication. Briefly, to summarize Bybee’s view:

a. Like “other types of stored mental percepts” (p. 37), linguistic representations are redundant, context-dependent, multiply specified, and procedural. For example, speakers know, in some sense, that of the three t’s in statistics, only the second gets aspirated.

b. Segmental units only “emerge from the inherent nature of the organization of gestures for articulation” (p. 85) So, for example, if the tongue is in the open [a] position, then the vocal cords need to vibrate if anything is to be heard; so we get consequently a recurring voiced [a].

c. Language change creates complementary distributions not only in phonetic but also in morphological contexts, so that the phonemic principle is not absolute (p. 213 et passim). For example, whether we make [ŋ] a phoneme in English will depend on whether we factor in morphology, as in finger and sing-er (if we ignore the morpheme boundary, then we need /ŋ/ as a phoneme).

6. Columbia School phonology without the phoneme

Like Bybee, Columbia School can do its work without the phoneme. To do so, we simply start from the communicative principle that signals must by and large (allowing for homophones) be made distinct. We can note, to whatever degree of detail we need, the phonetic characteristics that make one signal distinct from another. We can analyze these means of achieving distinctiveness and account for them just as we have always done (e.g., Diver 1979; Davis 1984 [1987]; Tobin 1997), but without going through the preliminary phonemic stage. We might, for example, use ordinary phonetic transcription without worrying about strict phonemic status. As an illustrative alternative, Figure 2 combines transcription with
matrices to describe the articulatory gestures comprising several lexical items of English. This is essentially the same idea as the “gestural scores” of articulatory phonology, where “gestures,” “the basic units of phonological contrast,” are “events that unfold during speech production,” consisting of “the formation and release of constrictions in the vocal tract” (Browman and Goldstein 1992: 155). Gestures overlap in time, producing “constellations” of gestures. Simply put, different articulators are doing different things at various times, and this is how distinctiveness is achieved.

Spaces between horizontal rows indicate relative aperture, from zero: -, =, =. Open space is larger aperture. Spellings and phonetic transcriptions are given for convenience.

<table>
<thead>
<tr>
<th>lip</th>
<th>shear</th>
<th>my</th>
<th>buy</th>
<th>pie</th>
<th>spy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
</tr>
<tr>
<td>apex</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
</tr>
<tr>
<td>dorsum</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>velum</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
</tr>
<tr>
<td>larynx</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
<td>[-]</td>
</tr>
</tbody>
</table>

**Figure 2.** A hypothetical graphic communication system

Each row of a matrix represents an articulator changing its state through time. Sheep is distinct from cheap in having initial aperture. Buy is distinct from my in having the nasal passage closed at the initial. Pie is distinct from buy in having late as opposed to early voice-onset. Spy is distinct from pie in beginning with an apical sibilant. Which pattern is preferred can be determined without preoccupying oneself over the phonemic status of affrication or aspiration. /č/ might well be a more complex unit than /š/; equivalently, cheap has an extra gesture lacking in sheep. Either way, sheep is preferred, and its type should be more frequent. /b/ might well involve an extra articulator, the glottis, over /p/. Equivalently, the late voice-onset of pie gives the speaker a bit of a rest with the vocal cords, with respect to buy. Either way, the pie type should be preferred. All this is to say that Diverian explanations for phonological favorings and disfavorings in the lexicon do not depend on the identification of phonemes.

Indeed, Diver’s phonology is more transparent if the phoneme is taken out. To see how, consider the four phonological explanations that Diver (1995) chose to use as his own illustration of how explanatory phonology is done:

Among languages, says Diver (1995: 70), an “extension’ along the axis of zero aperture, from /t/ to /p/ and /k/, is very common”; that is, most languages have not only a /t/ (which Diver speculates to have been historically “primary”) but also a
ADVANCES IN FUNCTIONAL LINGUISTICS

/p/ and a /k/. But these three phonemes are not equally exploited. Instead, “the greater investment in fine motor control in the case of the apex is reflected in a greater utilization of units formed with the apex, since its greater adroitness makes it easier to control.” Notice how the explanation is stated not in terms of phonemes /t, p, k/ but in terms of an articulator, the apex. That is because the explanation applies not only to /t/ but to apicals in general. Indeed, the statement might better read, “the greater investment in fine motor control in the case of the apex is reflected in a greater utilization of the apex,...” deleting the phrase “units formed with.” The generality of explanation continues: This skewing is especially noticeable, says Diver, where, “toward the end of the morpheme [where communicative load is lighter], apicals are progressively favored over labials and dorsals, sometimes to the exclusion of these two.” The explanation is stated in terms of articulators, not individual phonemes.

Diver (1995: 70) introduces his second illustration this way: “Beside /p t k/ there is often another group, using the same axes of articulation, /b d g/.” His explanation for the differential exploitation of these groups: “The coordination of an additional articulator, the glottis, requires more precision of control, and the more complex unit is regularly used with less frequency, in some languages down to zero.” Notice that the explanation applies uniformly across three phonemes and thus might better be stated without reference to phonemes: “The coordination of an additional articulator, the glottis, requires more precision of control, and so voicing is regularly used with less frequency. . . .”

The third illustration of explanatory phonology:

An entire chapter in a phonological study would be devoted to the co-occurrence of the phonological units within the morpheme. Co-occurrence is skewed rather than uniform. This is particularly noticeable in immediately adjacent units, where reuse of the same articulator is largely avoided, as in the case of tl-, dl-, pw-, sr-.

(Diver 1995: 70)

Notice how the phonemes are more or less tacked on as examples of the general fact. Why not simply say:

An entire chapter in a phonological study would be devoted to the reuse of articulators within signals. Reuse is skewed rather than uniform. This is particularly noticeable in the case of immediate reuse, which is largely avoided.

The last illustration of explanatory phonology presented in Diver (1995), and possibly the last he developed historically, lends itself especially obviously to a phoneme-free treatment.5 Referring to the prototypical architecture of signals built around a vocalic keystone with optional flanking constrictions – consonants – Diver observes that “looking outward from the keystone (as it were), uniformity
of usage of the vocal folds is preferred.” Stating this skewing in terms of phonemes would be awkward, little more than a recitation of the observations, and Diver himself avoids doing so. Perhaps the most succinct formulation in terms of phonemes might be: Initial /p/ and /t/ and /k/ prefer final /p/ or /t/ or /k/; initial /b/ and /d/ and /g/ prefer final /b/ or /d/ or /g/. Far more revealing to say, as Diver does, that, surrounding the keystone of the signal, “uniformity of usage of the vocal folds is preferred.” That formulation allows us to see that it is not a question of phonemes correlating with phonemes but of the articulation of voicing within signals. To paraphrase Diver’s formulation: Within signals, consistent voicing of constrictions is favored. Now we see more clearly the role of the human factor in this explanation.

It might seem curious that Diver, while retaining the phoneme, had little or nothing to say about allophones, those segments of sound which show partial phonetic similarity and which do not distinguish morphemes, since they appear in mutually exclusive environments. For example, the partially similar [k] sounds of keep, cop, coop, above, would be classed in descriptive linguistics as allophones of one phoneme /k/, each occurring in its particular pre-vocalic environment. While Diver limited himself to questions of the distribution of distinctive units, for example /k/ vs. /g/ in English, one could certainly pursue the question of why the various manifestations of /k/ appear where they do.

Tobin (1997: 21–23) provides a clear and succinct illustration for four allophones of /p/ in English, from fully aspirated to unreleased, as in pen, pajamas, spend, hip boot. The clear pattern is that /p/ is articulated with greatest energy (full aspiration) in the environment that is invested with greatest energy (stressed syllable), and with least energy (unreleased) where communicative load is lightest (word-final position). “In short,” says Tobin, “the synergetic relation between phonemes and their allophones may be seen as the primary example of the search for maximum communication with minimal effort,” that is, the interplay of Columbia School’s orientations of phonetics, communication, and the human factor.

If one were to do such phonology without appeal to phonemes and allophones, nothing of consequence would be lost. Tobin’s observations, like Diver’s, have crucially to do with phonetic facts (aspiration, stress), not with phonemes. Indeed, the observations are general across voiceless stops in English, not limited to any particular phoneme; /p/ serves merely as an illustration of the principle.

Another area investigated by Tobin (1997) is that of clinical applications, and here too explanations typically overarch the level of the phoneme. Tobin notes the debate within the clinical field over abstract phonology versus concrete phonetics, and he relates that to his and Diver’s distinction between explanation and description. Just as Diver used phonetics, communication, and the human factor to explain observations at the level of distinctive units, so those same explanations,
says Tobin, can be used to account for “speech disorders as representing favorings and disfavorings of certain allophones in different phonetic environments.” If this is so, the phoneme-allophone distinction introduces something unnecessary. The allophonic skewings are “similar to the favorings and disfavorings of certain collocations of phonemes in different phonetic environments [found] in the various languages examined by the theory of phonology as human behavior.” The explanations hold without regard to phonemic or allophonic status (Tobin 1997: 208).

It is difficult to imagine how dispensing with the phoneme would in any way substantively change the way Columbia School does phonology. Columbia School analyses have not depended crucially on the existence of phonemes but on voicing, visibility, aperture, adroitness, and so forth, characteristics which are not coextensive with individual segments. In truth, Columbia School phonology has never been a phonology of the phoneme but a phonology of the articulation of the signal, a phonology linking phonetics and communication, with the help of the human factor. In that enterprise, the phoneme was always superfluous.

Notes

1. While the spectrograph has confirmed that the phoneme cannot be a physical reality, its psychological status remains problematic. Eddington (n.d.) summarizes the difficulties.
2. Hockett (1987: 48) would much later affirm the validity of the “phonemic theory” but remain noncommittal as to whether it applied to phonemes, allophones, features, or something else.
4. Another explanation, not attributable to Diver, might be that the *pie* type maximizes distinctiveness between flanker and keystone (consonant and vowel) to a degree that the *buy* type, with its voicing throughout, does not; hence, *pie* promotes communication better than *buy*.
5. “The Phonology of the Extremes,” presentation at Rutgers University, Oct. 9, 1993; to be published in a volume being edited by Alan Huffman and Joseph Davis.

References


Chapter 9

Tell me about yourself

A unified account of English -self pronouns

Nancy Stern
The City College of New York

This paper offers an innovative Columbia School account of English -self pronouns (myself, yourself, etc.). The analysis rejects the view that the distribution of -self pronouns is a reflex of syntactic structure, as well as the traditional characterization of -self as a reflexive pronoun. Instead, -self forms are hypothesized to signal a constant meaning, insistence on a referent, which accounts for the forms’ distribution in authentic texts. This approach has led to the discovery that -self forms contribute to the same types of interpretations across a wide range of different structural contexts, including not only reflexive and emphatic uses, but also like-phrases, picture noun phrases, logophoric uses, conjoined expressions, and other environments.

1. Introduction

Traditional and modern treatments of English -self pronouns (myself, yourself, etc.) offer a classification of the uses of these forms. The best known of these uses are the reflexive and emphatic, but also included are a variety of other categories such as like-phrases and picture noun phrases. Some analyses (e.g., Zribi-Hertz 1989; Reinhart and Reuland 1993) also distinguish between what are called sentence-level and discourse-level uses. However, all these divisions – reflexive and emphatic, sentence and discourse – represent pre-analytical assumptions about the nature and distribution of -self. These a priori constructs have been deliberately avoided in this paper, which proposes instead that such categories are actually of limited analytical value, and that there are semantic commonalities which cross all these categories (Stern 2004a).

The goal of the research reported here has been to explain the distribution of -self pronouns – that is, to understand the linguistic motivation for the use of these
forms, and to account for the choices speakers evidently make between -self and simple pronouns (me, you, etc.). This paper begins with a brief review of the traditional account of -self pronouns. Next, the notion of reflexivity will be considered, and it will be shown that this concept does not contribute to our understanding the distribution of -self forms. Then, a different view of the data will be presented, through an examination of the types of messages with which -self forms are associated. That is, through analysis of contributions -self pronouns make to various communications, a different picture of their distribution will emerge. Lastly, I will propose a hypothesis about the meaning of -self forms, and explain how the proposed meaning, INSISTENCE ON A REFERENT, accounts for the wide variety of uses of these forms – including what are called reflexive and emphatic, as well as the so-called discourse uses.

2. **Traditional account of -self pronouns**

The traditional understanding of -self pronouns is based primarily on a structural distinction: reflexive uses are in head or argument position, while emphatics are adjuncts, or appositives.¹ The reflexive use is the cornerstone of this categorization. Traditional grammars note that reflexive pronouns are required whenever a noun phrase in a predicate is coreferent with the subject of the sentence, as in (1).

(1) a. Betty saw **herself**
   b. Barney talked to **himself**.

In the second major group of uses, emphatics, -self occurs in apposition, to provide emphasis, as in (2).

(2) a. Fred **himself** didn’t know the answer.
   b. I’ve never been there **myself**.

These structural categories form the basic schema for describing and analyzing the distribution of -self by formalists and functionalists alike. For the most part, research on these forms is divided neatly into studies of -self forms used as appositives and those used as arguments. There has been some interesting and illuminating work done on the semantics and pragmatics of what are called emphatics (e.g., Cohen 1999; Cresswell 1997; Edmondson and Plank 1978). In particular, I draw on the work of Kemmer (1995) and Kemmer and Barlow (1996), which, in Columbia School terms, focus on describing the messages for which appositive -self forms are used.
On the other hand, studies of *-self* pronouns in argument positions primarily aim to describe the structural conditions that obtain between the *-self* form and its antecedent. This type of research is exemplified by generative grammar’s Binding Theory, which holds that *-self* pronouns are a type of anaphor that must find their antecedents either within their own noun phrase or within their own clause. In the examples in (2), the *-self* forms refer to the entities named by the subjects of their clauses, and the referent of *-self* is mentioned elsewhere within the same clause.

In addition to these reflexive and emphatic uses, a long list of exceptions to this basic dichotomy is also recognized, as there are many cases in which the *-self* form finds no antecedent within its clause. The general outline of the traditional view of *-self* pronouns is schematized in Figure 1, which shows that the use of *-self* pronouns is first divided into emphatic and reflexive uses, followed by a list of exceptions that are then appended to the account.

This paper considers not just reflexive and emphatic uses, but also the apparent exceptions, including *like*-phrases and fixed expressions, picture noun phrases, logophoric uses, and conjoined expressions. Examples (3) through (10) below illustrate these types of exceptions.

Example (3) contains a *-self* pronoun without a clause-internal antecedent, in what is called a *like*-phrase. Example (4) is described as one of several *fixed expressions* in which *-self* forms may appear. Fixed expressions are those in which the *-self* pronoun is said to be optional; Quirk et. al. (1985) list the phrases *someone like -self, except for -self, as for -self*, and others.

(3) He would be something nondescript, something in the background, *like herself*; perhaps he had become an interpreter. (AMM)
… how gratifying it was not to have to explain or rationalize or defend her decision to anyone but herself. (GIN)

The next type of exception, shown in (5) and (6), is called a picture noun phrase. Picture nouns are the names of entities that represent likenesses of other entities, such as picture, story and photograph.

(5) [Regarding the author’s father:] I was visiting one time and he asked me to choose among several pictures of himself to send to “an old flame”. (SMF)

(6) “…He’s got this huge office, you know, and every square inch is covered with pictures of himself…” (EYB)

In (5), the referent of himself is not mentioned within its noun phrase (several pictures of himself), nor even within its clause (me to choose among several pictures of himself). Example (6) contains another picture noun phrase in which the referent of himself is not mentioned within its clause (every square inch is covered with pictures of himself).

Another group of exceptions, called logophoric uses, have also received a lot of attention from linguists. In these uses, the referent of the -self pronoun is the individual (or individuals) whose point of view is being represented, even though there is no overt clause-internal antecedent. The passage in (7) contains a logophoric use of -self:

(7) She felt a sudden rush of power, the power to stay alive. She’d kept others alive with her stories when they’d come close to being found. This time it was for herself. (HSR 383)

In (7), the referent of herself is a character named Trudi. There is no potential antecedent mentioned elsewhere in the clause, this time it was for herself, but it is Trudi’s point of view that is represented. Trudi has been arrested and is hoping to win her freedom by recounting stories to her guard. We learn her inner feelings (a sudden rush of power), and what she was thinking (she had told stories before). Example (8) is also logophoric:

(8) As he was about to slip the catch on the inside lock, Morrison glanced once more around the room…. Poor Louise had been trying to construct herself out of the other people she had met. Only from himself had she taken nothing; thinking of his chill interior, embryonic and blighted, he realized it had nothing for her to take. (APO 170)

Again, there is no clause internal antecedent: only from himself had she taken nothing; but it is Morrison, the referent of himself, whose point of view is represented.
Linguists of both functionalist and formalist stripe have treated logophoric uses as distinct in kind from other uses of -self. In a Cognitive Grammar account of -self pronouns, Van Hoek (1995) describes logophoric uses as “a more distant extension from the prototype” (p. 186), (the prototype being the reflexive). In generative accounts, Zribi-Hertz (1989) and Reinhart and Reuland (1993) have described logophoric uses as falling under the domain of discourse, rather than sentence-level, grammar. In their analysis, Reinhart and Reuland (1993) simply exclude non-reflexive uses of -self pronouns from the syntactic component of grammar. In what might be described as circular reasoning, the exceptions we have seen (i.e., like-phrases, picture noun phrases and logophoric uses) are excluded from the domain of the Binding Principles, and only reflexive uses are left as part of sentence-grammar.

-Self pronouns in conjoined expressions also occur without clause-internal antecedents, as in (9) and (10):

(9) There was too much of everything in the banker’s office, the banker’s house. It all only emphasized the gap between Elizabeth and himself. (Elizabeth is the wife of the referent of himself, and the banker is her wealthy father.) (HVE 31)

(10) I have a painting of Grandma, Priscilla, and herself, dressed for church, standing by the gate to our house in Holicong, and another of the view from our windows in Philadelphia. (MBW 69)

Many proposals have been made by modern linguists to accommodate the kinds of examples shown in (3) through (10). However, another conclusion to be drawn from this data will be explored in the following section: that the foundational schema – treating -self as a reflexive pronoun – is not analytically sound.

3. Reflecting on reflexivity

Attentive readers may have noticed that the popular appellation of -self forms, reflexive pronouns, has been assiduously avoided. This is, of course, quite deliberate, because using this term is not an analytically neutral practice. Rather, it reflects the assumption that the best description for these forms is that they are used reflexively, or that reflexive uses are in some way primary. While it is true that the majority of tokens of -self pronouns in discourse are reflexive uses, labeling and analyzing -self forms as reflexive pronouns actually obscures more than it illuminates.

If we accept the reflexive pronoun as an a priori category of linguistic analysis, then -self is probably the best candidate we can find in English. However, if we treat as a testable hypothesis the claim that -self is a reflexive pronoun, the hypothesis
fails. If -self really is a reflexive pronoun, we would expect to find -self pronouns in reflexive contexts, and simple pronouns, therefore, in non-reflexive contexts. Of course, we do find -self pronouns in reflexive contexts and simple pronouns in many non-reflexive ones. But we also find -self pronouns in non-reflexive environments, as we have just seen: as appositives, in like-phrases, in picture noun phrases, logophoric uses, and conjoined expressions, among others. We also find simple pronouns in reflexive environments, as shown in examples (11) through (13):

(11) He would have with him a bundle of food .... (LGI 160)
(12) Rankin locked the door behind her, and a guard was posted outside for her protection. (JAH)
(13) I lose control. It’s like, I’m not me. (OPR)

The examples in (11) through (13) all meet the classic definition of reflexivity: the pronoun refers to the same person as the subject. And yet, simple pronouns – not -self pronouns – appear. In Figure 2, I have adapted Diver’s (1987) approach to the Latin pronoun se to show the distribution of -self and simple pronouns with respect to reflexive environments in English:

<table>
<thead>
<tr>
<th></th>
<th>non-reflexive environments</th>
<th>reflexive environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>-self pronouns</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>simple pronouns</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**Figure 2.** Distribution of -self and simple pronouns (following Diver 1987)

The point of the chart in Figure 2 is this: reflexivity is not the conditioning factor that determines whether -self pronouns will occur. -Self and simple pronouns both occur in both reflexive and in non-reflexive contexts. Not only does reflexivity not explain the distribution of these forms, it doesn’t even describe them successfully. Rather, the notion of reflexive pronoun leads us away from the data, and away from a more successful analysis that can better account for the distribution of these forms.

4. **A new look at the data**

Once we shed the blinders of the a priori category reflexive, we are free to observe the data in a new light: namely, by looking at the communications to which these forms contribute, rather than at the structural conditions for their use. An exten-
sive examination of data culled from a variety of spoken and written texts has led to several new discoveries, including the observation that -self forms contribute to the same types of messages in both argument and appositive environments, as well as among what are called sentence- and discourse-based uses. This is a new view of the distribution of -self.

To understand the communicative import of -self forms, we will need to go beyond the very general label emphatic, to describe more precisely the message effects that result from the use of -self forms in apposition. Then, we will see that the same types of interpretations are found in argument environments as well. From a Columbia perspective, this is not surprising, for it suggests that -self pronouns are making the same semantic contribution in the wide variety of environments in which these forms appear.

The following description of the communicative import of -self pronouns is arranged according to various message types, but these message categories are simply a way of organizing a large amount of data. It is important to note that these rubrics are not proposed as analytical constructs or as discrete categories of any kind. The message effects are related to each other, and as we will see, most uses of -self actually exemplify more than one category.

4.1 Contrast/Comparison

One of the most common messages associated with -self forms is that of contrast or comparison with others. Examples (14) and (15) show appositive uses of -self that suggest a contrast between the referent of -self and others:

(14) Books about childcare … put so much emphasis on all the needs that children have … that parents sometimes feel physically and emotionally exhausted…. They get the impression that they are meant to have no needs themselves. (SBC 21–22)

The word themselves could have been left out of this passage, without changing what Columbia School calls the scene, or what other linguists call the truth conditions of the sentence. However, the presence of the -self form is an effective communicative resource that draws attention to the comparison – between the needs of the children, and those of the parents.

In example (15), the writer was better known by herself than by others:

(15) But she leaves behind as well a record of herself that is far more complete, more accurate, and more compelling than the lengthy chapter examining her life in the book Glittering Jewels of Japan, or in the various magazine articles about her that have appeared over the years. It seems that at least
Advances in Functional Linguistics

in the case of this one unusual subject, no one knew the memoirist as well as the memoirist herself. (GMG 2)

A comparison is suggested between the memoirist and others who have written about her.

Given this use of -self forms to communicate messages of comparison, it is not surprising that -self forms occur in like-phrases as in (3), repeated here as (16):

(16) He would be something nondescript, something in the background, like herself; perhaps he had become an interpreter. (AMM)

4.2 Exclusion

Another type of message found among -self forms in appositive uses is an exclusion of other entities. This interpretation is logically related to that of contrast, as the differentiation between the referent of -self and others is highlighted. In the next two examples, advertisers deploy -self forms to suggest that certain projects can be completed without the participation of others:

(17) You have to do your taxes, but you don’t have to do your taxes yourself! (HRB)

(18) Let JC Penney Custom Decorating measure, design, and install custom window treatments for you. [Or,] Measure for window coverings yourself…. here’s how. (JCP)

In both examples, the -self form is deployed to suggest that its referent can complete the actions alone. Other people who might also participate in these activities are understood to be excluded: that is, (17) could be paraphrased as “you don’t have to do your taxes alone”; (18) says something like “measure for window coverings without the help of our professionals”.

This type of comparison between the referent and other people, and the exclusion of others, can be seen in argument uses of -self as well, as in (4), repeated as (19):

(19) … how gratifying it was not to have to explain or rationalize or defend her decision to anyone but herself. (GIN)

There is a comparison between others (anyone) and the referent of herself; and the passage makes clear that a sharp distinction is being made between them. Anyone but herself should not simply be assigned to a general category called fixed expressions. Instead, it is an instantiation of a general tendency to use the meaning of -self to express an exclusion of others.
Concentrating attention to the exclusion of others appears to be a motivation for the use of \textit{-self} in (20) as well, about Jack Kevorkian, who had recently been jailed for assisting patients to commit suicide:

(20) But when they [young men during the Vietnam War] burned their draft cards, no one died. Their protest affected \textit{themselves} alone as sovereign individuals. (CKD)

Note that this example does not fit into the structural categories of reflexive, or emphatic, or into the list of exceptions we saw earlier. The use of \textit{-self} here appears to be motivated by the intention of zeroing in on the referent, specifically to the exclusion of others.

In (21), the speaker uses the word \textit{solely} twice, which also suggests that the referent of the \textit{-self} form is to be understood to the exclusion of others:

(21) “We created an ad campaign solely directed at women for this watch.” The thinking, she said, was that Patek Philippe needed “to be speaking to women in a language solely for \textit{themselves}.” (CAD)

Examples (14) through (21) share the semantic characteristic of highlighting a contrast between the referent of the \textit{-self} pronoun and others.

4.3 Unexpected Messages

Another frequent use of \textit{-self} forms as appositives is to serve as a “heads-up” warning to hearers, a notice that they should pay attention because there might be something about the referent that is not anticipated. In the next example, the referent of \textit{-self} is clearly unexpected:

(22) The second shock came when police identified the victims: Nicole Brown Simpson, 35, ex-wife of onetime football star O.J. Simpson, and Ronald Goldman, 25, a waiter and part-time model. Then came the almost incomprehensible news that the prime suspect in the case was none other than Simpson \textit{himself}.

The phrases \textit{the second shock}, \textit{almost incomprehensible news} and \textit{none other than} all contribute to the interpretation that the facts being described are indeed surprising and unexpected. Unexpected messages are one inference that may result from the appearance of \textit{-self} forms. We will return to this message effect in our discussion of reflexive uses in Section 4.6.
4.4 Importance of the referent

Another inference that may be derived from the use of -self is to add to the importance of any referent. Certainly repeated mention by appositives suggests greater importance in an iconic way, since the referent is worthy of being mentioned twice. In (22), the referent is a prominent, well-known celebrity, and the victim is identified in terms of her relationship to him (she was his wife). Appositive -self forms are frequently used for eminent referents, as in (23):

(23) Nell and Lillian stood at the front door and waved goodbye to Tom. Rachel wouldn't stir herself from a chair to say goodbye to Jesus Christ himself. (ABE 194)

In the next example, an argument -self form refers to entities who see themselves as quite important:

(24) Very young children are highly egocentric. The world revolves around them, and they see all events as focusing on themselves. (BPL)

This passage comes from the 4th edition of a textbook on second language acquisition, demonstrating that it is not an inadvertent use of the -self form, or some kind of performance error. Rather, the meaning of -self contributes to the intended communication here, part of which is to express the importance of the young children, at least in their own eyes.

4.5 Overlapping message effects

As we have noted, there is a great deal of overlap among message effects. The occurrence of -self in a particular context does not appear to be motivated by, nor an instantiation of, a single message category. Thus, in (22) it is not just the unexpectedness of the referent, but his importance as well, that motivates the occurrence of -self in this passage. In (23), the referent of himself, Jesus Christ, is – to say the least – very important, but he would also be quite an unexpected visitor. In (24), the children are important, but the meaning of -self forms is also used to zero in on those referents, to exclude others from the interpretation by suggesting that they see only themselves.

4.6 Role conflicts

Our communicative perspective can also account for the uses of -self that the traditional account describes as reflexives. We have seen that appositive -self forms are used for unexpected messages. I have proposed that reflexive uses are an instantia-
tion of this tendency, an observation also made by Kemmer (1995) and Levinson (1991). In reflexive environments, what is unexpected is that a single referent is playing more than one role at a single time. So, for instance, in the examples in (1), repeated as (25), Betty is both the seer and the seen – she has two roles in the act of seeing, and Barney has two roles in the talking – he talks, and is talked to.

    b. Barney talked to himself.

Thus, the meaning of -self forms may be used to suggest that referents are playing more than a single role at one time. This generalization, which I call the Role Conflict Hypothesis, accounts for the appearance of -self forms in reflexive environments, as well as in some environments that are not syntactically reflexive.⁵

The notion of role conflict is not simply a restatement of reflexivity. The role of the referent is not encoded by the -self pronoun, nor by a reflexive construction. Rather, role conflict is a possible inference resulting from a speaker’s use of a -self pronoun.⁶ And, importantly, the domain of role conflict extends beyond the boundaries defined by the concept of reflexivity. For instance, it applies to picture noun phrases and logophoric uses as well, which will be discussed in Section 5, in which a meaning hypothesis is also described.

5. A meaning hypothesis

Until now we have seen only a description of the distribution of -self forms, and no analysis has been presented. This has been done in an effort to highlight the new observations, and to separate them from the hypothesis that is proposed to account for them: the various uses of -self pronouns can all be explained by hypothesizing that -self forms are signals of meaning.

Like simple pronouns, -self forms signal Person, Number, and Sex meanings. But in addition, -self forms signal a meaning I have called INSISTENCE ON A REFERENT (Stern 2001). The postulation of such a meaning follows a well-established precedent in which linguists routinely assign meanings to forms, generally relying on a canon of well-established meanings such as past, definite, locative, etc. Although insistence is not a meaning that is sanctioned by familiarity (either within traditional accounts or even in previous Columbia School analyses), an examination and analysis of data can determine whether it is a successful hypothesis.

Insistence can be understood as a forceful pointing, an energetic reference that draws additional attention to an entity or entities. Thus, while rejecting the traditional characterization of -self as a reflexive pronoun, this analysis does borrow from the tradition by analyzing -self as a kind of emphatic pronoun. -Self
forms occur where their meaning, INSISTENCE, contributes to the intended communication. By INSISTING ON A REFERENT, speakers may communicate a variety of message elements.

For instance, we saw that -self forms contribute to messages in which others are excluded from reference. Excluding others is actually an effect of any reference, since any act of referring entails an exclusion of those who are not referred to. For example, in the act of referring to someone named Robin, those who are not Robin are not designated by that reference and so are necessarily excluded. Since INSISTENCE is a strong type of reference, the exclusion of others is concomitantly stronger as well. By INSISTING ON THE REFERENT in the ads about doing your taxes yourself, and measuring for window coverings yourself in examples (17) and (18), the -self forms stress the performance of the action by these referents, to the exclusion of other professionals who might also perform these tasks.

In (20) repeated as (26), the words alone and sovereign help the hearer infer that the speaker’s reason for INSISTING on these referents is to note only them, and to exclude other possible participants from the event.

(26) Their protests affected themselves alone as sovereign individuals.

By INSISTING on a referent, and thereby excluding others who are not included in the reference, -self forms may suggest a comparison between their referents and those who have been excluded. Messages of contrast or comparison are closely related to those of exclusion. After all, any reference includes an exclusion of those who are not referred to; an emphatic or INSISTENT reference entails a stronger exclusion, and therefore suggests a contrast between the referent, and those excluded by that reference.

Likewise, one INSISTS only on entities that are worthy of this extra energy; and, INSISTING confers additional attention, contributing further to the message that these referents are important. Unexpected situations provide another reason why a speaker may choose to INSIST ON A REFERENT. By INSISTING, a speaker draws additional attention, so the meaning is well-suited to messages in which the referent’s identity will be surprising to hearers.

We noted earlier that picture noun phrases and logophoric uses are anomalous in structural accounts of -self pronouns. They are banished from the domain of sentence grammar, and driven to the realm of discourse grammar. But from a semantic perspective, these types of messages are fully consistent with our meaning hypothesis.
5.1 Picture noun phrases

Picture nouns suggest the existence of a viewer, or perceiver, of the likeness they represent. Thus, there are at least two roles suggested by any picture noun: the entity portrayed, and the entity perceiving. In example (27), seen earlier as (6), Donald Trump has more than a single role with respect to the picture noun. This time, more context is included:

(27) “...Donald wants to sit down with me. So I fly up to New York. What the hell, right? And I walk in his office and I just crack up. He's got this huge office, you know, and every square inch is covered with pictures of himself. Like a shrine to Donald, you know? He starts telling me how he's thinking about doing Trump.com, selling Trump this and Trump that, perfume and other stuff, and then he's going to take it public.” (EYB)

Trump has multiple roles here: he owns the office, he has photos inside that office, and he is pictured in the photos. As noted earlier though, it is not just one message effect that motivates the use of a -self pronoun. In this example, Trump has summoned the writer to his shrine-like office, where he talks about his grand plans for Trump-this and Trump-that. Another motivation for insisting on this powerful referent is surely to take note of his importance and grandiosity.

5.2 Logophoric expressions

Logophoric -self forms are motivated by and contribute to many of the same message types we have seen. By definition, the referent of a pronoun used for a logophoric message plays at least two roles in an utterance: the role for which he is mentioned, and as the cognizer of the situation. Zribi-Hertz (1989) calls this role the “Subject of Consciousness” or “SC”.

We have seen that -self forms are used when there is something unusual or unexpected about their referents, and in logophoric uses, there are at least two elements that are unusual: the referent of the form has more than a single role, and the perspective taken is not that of the speaker. The normal expectation in any discourse is that a speaker offers her own point of view. An utterance such as Trudi felt a sudden rush of power is the author’s point of view of what the character felt. If a speaker wishes to deviate from that typical situation by expressing the point of view of a particular character, that is surely unusual and noteworthy. We have also seen that only entities with a certain measure of prominence are referred to by -self pronouns. A character whose point of view usurps that of the author is clearly important.

Let’s look specifically at the logophoric examples we have seen, repeated as (28) and (29).
(28) She felt a sudden rush of power, the power to stay alive. She’d kept others alive with her stories when they’d come close to being found. This time it was for herself. (HSR 383)

(29) As he was about to slip the catch on the inside lock, Morrison glanced once more around the room…. Poor Louise had been trying to construct herself out of the other people she had met. Only from himself had she taken nothing; thinking of his chill interior, embryonic and blighted, he realized it had nothing for her to take. (APO 170)

Throughout the story in which (28) appears, the author is addressing the reader, narrating the story from an external perspective. The unexpected use of herself alerts the reader that there is something unusual about her: the thought (this time it was for herself) is not from the perspective of the narrator, but from Trudi’s own point of view. In addition to the fact that the stories were for her, she is the cognizer of the situation, the Subject of Consciousness. There is also a conceptual role conflict, in that her storytelling was for herself. In the passage, the word it refers to Trudi’s storytelling, so she has two roles in the interpreted utterance: she is telling stories for herself. In addition, the meaning INSISTENCE contributes to a comparison between her telling stories for others, and for herself.

In (29), the use of himself also suggests to the reader that there is something unexpected, and in this case, it is that Morrison has two roles – he is the one thinking that she took nothing from him. -Self gives the sense of an indirect quote, that the reader is witnessing Morrison’s own thoughts. But once again, this writer has not chosen just the message category logophoric. As the word only suggests (only from himself), the meaning INSISTENCE concentrates attention on this referent to the exclusion of others, another motivation we have seen for the deployment of the meaning INSISTENCE.

5.3 Conjoined expressions

There are many reasons why one might INSIST on a referent in a conjoined expression, just as we have seen there are many reasons one might INSIST on a referent in another context. However, one of the characteristics of conjoined expressions makes the meaning INSISTENCE especially useful: the word and, which draws things together, creating the notion of a group consisting of the entities named by the two conjuncts.8 -Self forms may indicate that their referents are special, noteworthy members of the group formed by the conjoined phrase, or an unexpected role or role conflict for their referents.
In example (30), seen earlier as (9), the *-self* form contains no clause-internal antecedent, yet it is entirely felicitous as it contributes to the writer’s message that there is an enormous gap between the referent of *himself*, and his wife:

(30) There was too much of everything in the banker’s office, the banker’s house. It all only emphasized the gap between Elizabeth and *himself*. (HVE 31)

Again here, the use of the *-self* form is multi-determined. This passage is logophoric, since it is the referent of *himself* who is in the office of the banker (his father-in-law), and who is thinking that the banker’s lifestyle is further evidence of the great gulf that exists between himself and his wife.

Example (10) is repeated here as (31), with additional context. Here there is no logophoricity, as the narrator of this portion of the text is not the referent of *herself*. Nevertheless, the *-self* form appears in a conjoined expression, without a clause-internal antecedent:

(31) Only Elizabeth was an artist to her fingertips, and whichever of her talents she used, she used it differently from the rest of us. Her perceptions, so different from ours, have nourished me through the years. Her understanding of what has gone on in schools has provided depth and life to my own observations on American education. And her paintings have made every place I have lived my home. I have the first watercolor she ever painted and a painting of the New York Stock Exchange building she made for father. I have a painting of Grandma, Priscilla, and *herself*, dressed for church, standing by the gate to our house in Holicong…. (MBW 69)

The phrase *a painting of Grandma, Priscilla, and herself* can be described as both a conjoined expression and a picture noun phrase. What is noteworthy is that the same factors motivate the use of the meaning INSISTENCE in this example as in others we have seen. The expanded context shown in (31) shows that the referent of *herself* has more than one role with respect to the painting, as she is not only a subject in the painting, but also its creator. For this reason, she should be differentiated from the other members of the group formed by the conjuncts, as she is a special member of this group. The referent’s importance to the writer is another factor motivating the choice of the meaning INSISTENCE here. Not only is Elizabeth the topic of this paragraph, but the love and respect the author feels for her are expressed in this passage as well.

As we have seen, there is a great deal of overlap among these message effects. It should be noted that there also appear to be point-of-view effects in several other examples we have considered, such as the *like*-phrase and *anyone but herself* examples.
6. Summary

None of the message effects described here is present in every use of -self. Nevertheless, we have seen that the same message effects that are found among appositive uses of -self forms are also found among arguments, including reflexives as well as so-called discourse uses, including picture noun phrases and logophoric uses. This is a striking finding, and it strongly suggests that -self forms are making the same semantic contribution in all these environments. What once appeared to be a rather ad hoc assemblage of reflexives, emphatics, and a rag-tag list of other cases has now been shown instead to have common semantic characteristics.

What is common to every use of -self – appositive and argument uses alike – is the meaning INSISTENCE. This Columbia School semantic analysis of -self forms explains a variety of seemingly disparate uses of -self in a unified and theoretically coherent way.

Acknowledgements

I would like to express my deep appreciation to Ricardo Otheguy for his generous support and encouragement of the research on which this paper is based, and to the participants at Columbia seminars, Joseph Davis, Radmila Gorup, Alan Huffman, Wallis Reid, and Betsy Rodríguez-Bachiller, for their valuable assistance as well.

Notes

1. The term argument is used here in opposition to appositives – to refer not only to arguments selected by the verb but to all mentions of entities in head position. So, in the following utterance, the names of Wilma, picture, Fred and newspaper are all described as arguments: Wilma was looking at a picture of Fred in the newspaper.

2. See Stern (2001) for a more thorough analysis of all these uses.

3. Examples like (13), technically a predicate nominative, typically contain a -self pronoun, i.e., I'm not myself.

4. Another analytical possibility is that the definition of reflexive should be changed (cf. Reinhardt and Reuland 1993). However, such a strategy would rest on the presupposition that reflexivity is still the key to understanding -self pronouns, and that this a priori category should be salvaged, in spite of empirical evidence that does not support it.

5. It also explains the occurrence of simple pronouns in grammatically reflexive environments (Stern 2001).

6. The term speakers is used here interchangeably for both speakers and writers.
7. The exclusion is not limited to those with a different name. Clearly, other people who are also named Robin are also excluded.

8. See Stern (2004b) for a more thorough discussion of -self pronouns in conjoined expressions.

Data Sources

ABE  

AMM  

APO  

BPL  

CAD  

CKD  

EYB  

FKG  

GIN  
George Gendron. “The great achievement of the past few decades is that we now know everything necessary to grow a business,” Inc. May 18, 1999. [InfoTrac*]

GMG  

HRB  
H&R Block TV Commercial, March and April 2000.

HSR  

HVE  

JAH  
Glen Jeansonne. “The Lone Dissenting Voice,” *American History*, April 1999 (v34 i1 p46(6)). [InfoTrac*]

JCP  

LGI  

MBW  
OPR  An audience member on the television show, “Oprah.” ABC July 1, 2000


*InfoTrac  The InfoTrac Reference Center, Magazine Index from the General Reference Center. 2001. The Gale Group Inc. This electronic database consists of articles from magazines, reference books, newspapers and news services on current events, popular culture, the arts and sciences, sports, business etc. that have appeared from 1980 to the present.

References


Arguing that the distribution of *se* in Serbo-Croatian cannot be explained by invoking *a priori* categories reflexive, impersonal, middle voice, etc., this analysis follows García (1983) and Davis (2000) to advance a hypothesis that *se* is a signal in the semantic substance of Participant Focus. Whereas other pronouns in this complex system signal a variety of meanings (in several semantic systems), *se* says that there is a bona fide participant centrally associated in the event named by the verb but does not give any more information about it. This study claims that the opposition of substance between *se* and all the other Participant Focus forms accounts for its distribution in Serbo-Croatian.

1. **Introduction**

Serbo-Croatian *se*, the short form of *sebe*, has traditionally been called the reflexive and impersonal pronoun. García (1975) and Diver (1987), rejecting traditional syntactic categories, analyzed the Romance cognate of *se* as a signal of relative meaning, low deixis, arguing that the referent of *se*, whether true reflexive or impersonal, requires or deserves less attention from the reader than that of other pronouns in order to be identified. The present analysis of Serbo-Croatian *se* follows instead García's (1983) and Davis's (2000, 2002) analyses dispensing with the semantic substance of Deixis in favor of an opposition of substance. That is, *se*'s distribution is accounted for by the fact that *se* signals less information than other pronouns. This short morpheme, the fourth most frequent word in Serbo-Croatian, offers a wealth of insights into the working of the language (Kostić).

Both *se* and *sebe* are unmarked for person, gender and number. The full form *sebe* is marked for case. In addition to *sebe*, the accusative form of which *se* is the clitic, *sebe* has only the full form in the genitive *sebe*, the dative/locative, *sebi*, and the instrumental, *sobom*. In Croatian, *sebi* has the clitic *si*.2
In the literature, the number of homophonous se’s postulated by different analysts varies from one to ten. The basis for these classifications is sometimes semantic, sometimes syntactic, or at other times, a combination of the two. In her (1988) study, Đorđević distinguishes ten functions of se. The first five are reflexives and the other five are combinations of middle voice and passive. The author admits that the boundaries between these categories are not always clear-cut and that they often overlap, so she avoids the analytical problem by simply calling se a “multifunctional grammatical device”. Unfortunately, this could be said for most, if not all, grammatical forms.

Se is sometimes referred to as an empty morph or a verbal particle, but most often grammarians call it a reflexive pronoun. However, se is not always reflexive. In example (1) se does not refer to the subject.

(1) Kaže se da vreme leči sve rane.
   says (3sg.) se that time heals (3sg.) all wounds.
   ‘It is said that time heals all wounds.’

Grammarians classify examples such as this as impersonal usages. While it is true that we can often infer impersonal messages from examples in which se occurs, we should note that impersonal messages can also be signaled in a variety of ways in Serbo-Croatian. Examples (2) to (4) will illustrate this.

(2) Čovek nikad ne zna na čemu je.
   man never not knows on what is
   ‘One never knows what he can count on.’

(3) Treba pročitati ceo tekst da bi se razumeo njegov smisao.
   needs (3sg.) to read entire text to understand its (masc.) sense
   ‘One needs to read the entire text to understand its meaning.’

(4) Najveći mogući specijaliteti Pariza su plodovi mora. Na metalnoj konstrukciji donesu ostige, razne vrste školjki rakove, puževe. U podnožju stave puter, limun, sos od sirčeta i belog luka…. (PK 130)
   ‘The greatest possible speciality of Paris is seafood. On a metal structure they bring oysters, various shells, shrimps, snails. At the bottom they place butter, lemon, a sauce of vinegar and garlic…’

In (2), čovek (nominative case, singular), in (3) treba (3sg.), and in (4) donesu (3pl.) and stave (3pl.) are variously used to refer to anyone and everyone. All these examples express impersonal messages.
The labels reflexive and impersonal, generally used by grammarians for se, as well as the terms empty morph or verbal particle, have no explanatory power: they do not help us to understand what motivates the actual distribution of se. Once a large corpus is examined it becomes clear that the distribution of se does not correspond to these categories.

To account for the distribution of se this paper proposes a single meaning for the form, a common semantic input that se contributes to the message in every instance in which it occurs. Before we turn to a description of the proposed meaning of se and its relation to other meanings in the pronominal system, it is necessary to note that, in Columbia School, in addition to the grammatical meanings of the kind proposed here, the human factor (e.g., intelligence and inferential powers) also plays a major role in the use and shaping of language categories. While recognizing that linguistic signs consist of forms and their meanings, we also recognize that meanings are not precisely articulated and encoded in the language so as to allow a straightforward mapping in which the linguistic input corresponds directly to communicative output. Rather, signaled meanings provide only hints: it is up to a reader or hearer to take the hints provided along with other information available in the communicative act to bridge the conceptual gap between the meanings of the linguistic forms and any messages to be inferred.

In this paper, the hypothesis (following Davis 2000, 2002) is that se is a signal in the semantic substance called Participant Focus. The meanings of Participant Focus indicate how central or peripheral an entity is in the event expressed by the verb. Se signals that there is a genuine participant in central Focus. In other words, se signals that there is a participant centrally associated in the event named by the verb. Figure 1 sets out the proposed meaning of se.

Se has a rather sparse meaning when compared to the meanings of the other signals in this complex system. On the one hand, se is opposed to the other forms which also signal the meaning central Focus but which in addition to Focus signal a variety of other meanings such as Number, Gender and Discourse Referent. For example, ja ’I’ signals one, speaker, and central Focus. Se signals only central Focus.

Se is opposed as well to the peripheral Focus signals, which in addition to information about Number, Gender and Discourse Referent signal Control. Se does not signal any of these meanings. Thus, while ga ’him, it’ signals one, peripheral Focus, and low Control (masculine gender), se just signals central Focus.

This study advances the claim that the opposition of substance between se and all the other Participant Focus forms accounts for its distribution in Serbo-Croatian.
2. **Neutralization of the substances Number, Gender and Discourse Referent**

It is precisely the paucity of its semantic input that makes *se* a very useful tool of communication. The most interesting properties of *se* derive not from what it signals but from what it does not signal. The very imprecision of *se* allows it to function as a versatile instrument of communication.
Since *se* gives no information about Number, Gender and Discourse Referent, from messages in which *se* occurs one may infer that this information is not necessary in that particular instance of communication:

(5)  *Znalo se* da su preci poslednjeg konta Fanfonja postepeno gubili svoje posede. (PR83)

‘*It was known* (knew *se*) that the ancestors of the last count Fanfonja have been gradually losing their possessions.

Here *se* tells the reader that there is a participant in *znalo* but that it need not be individualized. *Se* can apply to all: masculine, feminine, and neuter; one and many; speaker, hearer, or other. *Se* only says that the participant of *knowing* is the main (central Focus) participant but remains indifferent to Number, Gender and Discourse Referent. *Se* here refers to no particular person. Without *se*, the ending of the finite verb would tell the reader that there was a specific referent. In example (5) with *se*, the effect is a neutralization of the substances of Number, Gender and Discourse Referent.

Because *se* does not signal Number, Gender, or Discourse Referent, utterances with *se* may lead the hearer/reader to infer that the candidate for central Participant Focus is a nameless situation rather than a specific entity. In (6), the verb *desilo* ‘happened’ does not refer to a specific entity:

(6)  *Početak starenja svog tela doživela sam kroz promenu svojih telesnih mirisa. Desilo se* to odjednom oko 33. godine moga života. (PK115)

‘I experienced the beginning of the aging of my body through the changes of my body odors. *It happened* (*se*) suddenly when I was 33 years old.’

In examples such as this, *se* blocks the inference that there is a referent that can be defined in terms of Number, Gender or Discourse Referent. Because there is no such information signaled, the reader does not seek a referent to match such a description. Without *se*, the finite verb would be taken to refer to a specific referent.

Whereas in example (6) the information about Number, Gender, and Discourse Referent is irrelevant, in other cases it may be redundant because the referent of *se* is obvious. In such a case, explicit information about Number or Gender might confuse the reader by suggesting that he try to find some participant that matches the description. In (7) no information is given about Number, Gender, or Discourse Referent:

(7)  *Okrenuo sam se* i veselo joj mahnuo rukom. (PR23)

‘I turned (*se*) and happily waved (with my hand) to her.’
Since *se* provides no information about Number, Gender, or Discourse Referent, the reader need not make an effort to look for another participant, but rather infers that there is only one participant (‘I’) who turned. Note there is no distinction being made between agent and patient.

Traditionally, examples such as this are called *true reflexives*. However, statistically these examples represent a relatively small number of tokens in the language. Much more numerous and also much more interesting are examples in which the referent of *se* is the only participant in the event and in which *se* interferes with the participant ranking established by the system of Control.

3. Neutralization of Control

As seen in Figure 1, Serbo-Croatian has a three-level Control system (using clitic pronouns): **mid** Control is signaled by the meanings of the dative pronoun clitics (*mi, ti, mu, joj, nam, vam, im*); and **low** Control is signaled by the meanings of the accusative pronoun clitics (*me, te, ga, je/ju, nas, vas, ih*). **Mid** and **low** are the only meanings with dedicated signals of control among the clitic pronouns. High Control is generally inferred. We can use the terms **one-**, **two-**, and **three-participant role** to describe the number of Controlling participants signaled or inferred for a particular event. However, Control, which is not the subject of this analysis, may also be inferred from other lexical items and other pronouns.

In all the examples in which the referent of *se* plays only a one-participant role, *se* occurring with an explicitly mentioned entity allows the inference that its referent is playing a Control role other than expected. In order to illustrate the effect *se* has on one-participant messages, two examples with the events of *proneti* ‘to carry through, to spread’ (news or rumors) are compared, first in (8a) with two controllers and then in (8b) with *se*:

(8) a. Ugljar Sava *pronese* glas da to kovač pravi mašinu koja sama od sebe leti. (MV 198)

“The stoker Sava *spread* the news that the blacksmith is making a machine which flies by itself.”

b. *Pronela* se vest o radnji nekog trgovca Đovanija… (PR128)

“The news *spread* *(se)* about the store of certain merchant Giovanni…’

The verb *proneti* ‘to carry through, to spread’ implies another participant in addition to the high Controller. In (8a) two participants are mentioned, ‘the stoker’ and ‘the news’. Quite naturally, the participant in central Focus, ‘the stoker’, is
Chapter 10. Se Without deixis

inferred to also be the agent of spreading the news (high Control). ‘The news’ is signaled by accusative case to exercise low Control over the event of ‘spreading’ – a patient role. Two distinct participants are thereby ranked in Control of ‘spreading’: one is inferred to be in high and one signaled to be low Control.

By contrast, no ranking of Control is necessary in (8b). There is only one participant and it is an obvious (the only explicitly mentioned) candidate for Control over the ‘spreading’. However, the presence of se in (8b) suggests that the referent of se, even though a participant in central Focus, is not in control of the event. News cannot spread by itself: someone spread it, but that someone is not specified here. Rather, the patient in the event of ‘spreading’ is put in central Focus and se alerts the reader that the participant in Focus is not in control of the event. Examples such as (8b) in which the grammatical subject plays the role of patient are the most frequent and widespread examples with se.

Consider now examples with the verbs otvoriti and rastvoritii both glossed ‘to open’. Examples (9a) and (9b), transitive and with accusatives, come first. They are followed by examples with se (9c) and (9d):

(9)  
a. Napokon su vojnici otvorili težak poklopac na kupoli tenka. (PR 76)  
‘Finally, the soldiers opened the heavy cover at the top of the tank.’

b. Vrata su otvarala put ka stepeništu koje je vodilo nadole u podrum i nagore etažima. (PV 38)  
‘The door opened the way to the staircase leading down to the cellar and up to the floors.’

c. Nije, koliko sam mogao da vidim, bilo povređenih…. Izašao sam iz kola i pošao ka prvoj karici lanca. Vrata su se otvorila dok sam još prilazio… Starac u džemperu sa motivom jelke i bundom koja je se-zala do članaka iskoračio je iz zadimljene unutrašnjosti. (OD 52)  
‘As far as I could see there were no injuries… I got out of my car and headed toward the first link of the chain (car in a chain accident). The door of the car opened (se) while I was still approaching…An old man wearing a sweater decorated with a Christmas tree and a fur coat reaching to his ankles came out of the foggy interior.’

d. Crkvene se dveri, tvrdo gradiene od hrastovine i kovanog gvozđa, na ovaj znak rastvoriše. (OCSS 10)  
‘At this sign, the church doors, solidly built from oak and wrought iron, opened (se).’

In (9a) and (9b), there are again two distinct participants: soldiers and cover in (9a) and door and way in (9b). In both examples, one participant is inferred to be a high
Controller (*soldiers* and the *door*) and the other as a low Controller (*the cover* and *the way*). Two distinct participants are ranked in the event of opening.

In (9c) and (9d) in which *se* occurs and only one participant is mentioned, there is no participant ranking. Example (9c) is part of a description of a chain of collisions. The narrator comes out of his car and heads toward the first car involved in the accident in order to see if anyone is hurt. Focus is clearly on the door of the car rather than on the driver. The narrator can’t even see the driver because of the foggy interior. In (9d), in which *rastvoriti*, a synonym of *otvoriti* occurs, the narrator might have said: “The monks opened the church doors.” That is what most probably happened. However, the narrator chooses not to focus on the monks.

In the scene in (9d), a religious procession is circling the church. Once the third circle is completed, the chief priest makes the sign of the cross in front of the doors. The doors are then supposed to open. At this point, all eyes are focused on the doors. Focusing on the monks here would be inconsistent with the intended message. Instead, *the doors* are grammatically in central Focus with respect to this opening. Of course, they are not the agents, the high controllers, of the opening. *The doors* are the thing opened, not the opener. In this example *se* blocks the assignment of two distinct Control roles. The presence of *se* here suggests that where there might have been two distinct participants – the door and the opener – only one is mentioned. Because *se* does not signal Number, Gender, or Discourse Referent meanings, it will not be construed to refer to an additional participant. There is no need to rank participants.

The following two pairs of examples further illustrate the neutralization of the System of Control which motivates the use of *se*:

(10) a. Naše novine su škro *javile* da je Vlada otišao u Italiju gde je počelo snimanje filma “Taras Bulja”. (PR 48)

‘Our newspapers briefly reported that Vlada left for Italy where the filming of the movie Taras Bulba has started.’

b. *Javljali su se prvi izlivi* nekog skrivenog besa, čudna mešavina jarosti i nemoći. (PR 29)

‘There appeared (se) the first outpourings of a certain hidden fury and a strong mixture of rage and powerlessness.’

(11) a. Ne znam da li je *ovu moju blagonaklonu opomenu čula jedna žena* u prolazu. (PR 95)

‘I do not know whether the woman who was passing by heard this well intended warning of mine.’
b. Bila je prilika da se čujo razne neobične priče, a najčešće one koje su se odnosile na naše doživljaje stećene u ratu. (PR 105)

‘This was the opportunity to hear (se) strange and varied stories (for strange and varied stories to be heard) most of them referring to our experiences lived during the war.’

In (10a) and (11a) two participants are ranked in the events of ‘reporting’ and ‘hearing’. In examples (10b) and (11b), the entities in central Focus, ‘outpourings’ and ‘stories’, are by their nature unlikely to be high Controllers. The presence of se hints to the reader not to consider the participant in central Focus to be an agent. Even though the participants are centrally involved in the event, they are not in high Control.

In examples (9c), (10b) and (11b), the entities in central Focus are inanimate objects or notions which ordinarily do not play the role of agent. In the next example, the participant in central Focus is a human and as such ordinarily deserves to be considered for that role:

(12) Peva se on i drugde i u crkvama u Skoplju ali se posebno radosno razlaže ovim hramom. (OCSS 11)

‘He is sung (se) elsewhere, also in churches in Skoplje, but he resounds (se) especially joyfully in this temple.’

This example is rather unusual and needs more context. The scene is the entrance of a procession into a church. The guest of honor, the king’s priest, is entrusted with bringing back to the king the Pascal canon of St. John Damascene, a saint who composed beautiful church music. The narrator makes an important comment saying that while the saint’s songs are sung elsewhere, they sound particularly beautiful in this church. The se in this example signals that the referent of se (on ‘he’) St. John Damascene, is the participant in central Focus. Still, in the events of ‘singing’ and ‘resounding’ the saint has a diminished (non-controlling) role. The saint does not do the actual singing; his songs are sung by others and their singing resounds well in this particular church. Even though the saint is designated as the main participant and put in central Focus in this singing, he is not in control.

4. Human referent of se and inferring Control

While se does not signal any degree of Control per se, se does affect the control ranking system in a significant way. Se only says that there is no control ranking in these messages. What roles are to be inferred depends very much on the meaning
of the verb, the characteristics of the referent of *se* and the surrounding context. In examples in which the referent of *se* is an animate entity, and thus a candidate for a role of high Controller, *se* often allows the inference that the participant in Focus is affected by an outside force. Examples (13) to (16) illustrate this:

(13)  
Ja sam *se* i dalje čvrsto *držao* svog iznenadenja: Tito u kontejneru! (PR 93)  
*I remained firmly gripped (se) in my surprise: Tito in a garbage bin!*

(14)  
Došavši u Trst, ona *se* zbušeno *našla* u jednom stranom gradu u koji je došla sa premalo para. (PR 128)  
*‘She came to Trieste and to her embarrassment found herself (se) in a foreign city to which she came with too little money.’*

(15)  
Jednom sam o tome razgovarala sa nekim piscem... Zanima me kako bi on opisao taj trenutak kada *se otvaram* prema drugom i kada prestajem da budem ja, kada ego isčezne negde izvan sebe. (SDK 8)  
*‘I spoke once about that with a writer… I was interested to know how he would describe the moment when I open (se) (I am opened) to the other and when I cease to be I, when ego disappears somewhere outside itself.’*

(16)  
U poodmakloj trudnoći žalila *se* na gorušicu i gasove. (UP 17)  
*‘Late in her pregnancy she complained (se) of heartburn and gas.’*

In all these examples *se* interferes with the Control ranking in the above messages. The presence of *se* hints to the reader that its referent is playing a role on the Control scale other than the expected. From the meaning of *se* in combination with the surrounding context, the reader infers that the participant in central Focus is affected by an outside force. For a variety of reasons the participant has a diminished capacity to act or displays a lack of control in various ways.

What is exactly affecting the participant in Focus is usually introduced in the context. In (13), with the verb of ‘gripping’, what causes the lack of control is not a physical thing but rather the shock of seeing the collected works of Tito in a garbage bin. In example (14), an embarrassing situation causes the lack of control. A young woman comes to Trieste specifically to shop. She feels uneasy not only because she is in an unfamiliar environment but more so because she realizes she does not have enough money to do her shopping. The expression ‘to her embarrassment’ supports the inference that the participant lacks control. In (15), a young woman explains the overwhelming effect of a passionate kiss. She is so affected that she feels totally lost. In (16), the participant in Focus feels uncomfortable in late pregnancy. She pities herself, feeling like a victim.
With verbs that are likely to suggest two participants, such as *držati* ‘to hold’, *naći* ‘to find’, *žaliti* ‘to pity’, *otvoriti* ‘to open’, etc., *se* subverts the Control ranking. Instead of a Control signal we have *se*, which tells the hearer/reader not to look for a second controller. The signaled participant does play a role, but not what one might expect. Some outside force is causing lack of control. What affects the participant in Focus is often given in the surrounding context. In these examples, entities which are ordinarily taken to be agents are affected by forces beyond their control.

There are verbs in Serbo-Croatian which overwhelmingly occur with *se*, many of which are called *middle voice* by grammarians. From the lexical meanings of these verbs and the presence of *se* the reader can infer that the participant in Focus is not performing a usual High Control role; this is often because the participant has been affected by an outside force as in examples (13) to (16). A partial list of these verbs is found in Table 1:

<table>
<thead>
<tr>
<th>Verb</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>truditi se</em></td>
<td>‘to make an effort’</td>
</tr>
<tr>
<td><em>diviti se</em></td>
<td>‘to admire’</td>
</tr>
<tr>
<td><em>snaći se</em></td>
<td>‘to cope with, to find one’s way’</td>
</tr>
<tr>
<td><em>boriti se</em></td>
<td>‘to fight’</td>
</tr>
<tr>
<td><em>bojati se</em></td>
<td>‘to be afraid’</td>
</tr>
<tr>
<td><em>nadati se</em></td>
<td>‘to hope’</td>
</tr>
<tr>
<td><em>odupreti se</em></td>
<td>‘to resist’</td>
</tr>
<tr>
<td><em>odreći se</em></td>
<td>‘to renounce, to waive’</td>
</tr>
<tr>
<td><em>snebivati se</em></td>
<td>‘to hesitate’</td>
</tr>
<tr>
<td><em>stideti se</em></td>
<td>‘to feel shame’</td>
</tr>
<tr>
<td><em>desiti se</em></td>
<td>‘to happen’</td>
</tr>
<tr>
<td><em>obreti se</em></td>
<td>‘to find oneself’</td>
</tr>
<tr>
<td><em>osvrtati se</em></td>
<td>‘to look back’</td>
</tr>
<tr>
<td><em>obazirati se</em></td>
<td>‘to look over one’s shoulder, etc.’</td>
</tr>
</tbody>
</table>

There is something in the lexical meaning of these verbs that suggests that even a participant in Central Focus, though generally a good candidate for an agent role, in this kind of event will have diminished capacity and thus less control (e.g. shame, fear, or hesitation). However, *middle voice* is not a category in the language with a dedicated morphology of its own. Rather, it is *se* which, together with the lexical meanings of these verbs, blocks the inference that a participant is an agent with a high degree of Control. 6
5. **Two-participant messages with se**

In the examples which follow, (17) – (19), in addition to one explicitly mentioned participant put in **central Focus** by the verb-ending and se, we have another participant which might ordinarily be inferred to be a high Controller. However, here that participant is signaled to be **mid Controller** (mi, ti, mu, etc., as shown in Figure 1). Though the mid Controller, most often a clitic, still possesses a relatively high degree of control, in these examples there is always an outside factor, a **super-agent** that is causing a certain effect or impression.

(17) *Lice mi se iskrivilo u grimasu čuđenja.* (NS 21)

face me (Dat.) se distorted into a grimace of wonder

‘My face distorted (se) into a grimace of wonder.’

(18) *Ali još za vreme jela oči su im se sklapale od umora i samo su sa krajnjim naporom uspevale da izgledaju kao da slušaju.* (D 163)

‘But already during the meal their eyes were closing (eyes them (Dat.) se closed from tiredness) from tiredness and only with the utmost effort they succeeded to pretend to listen.’

(19) *Noge mi se ukočiše.* (PK 112)

legs me (Dat.) se got stiff

‘My legs got numb (stiff) (se).’

Most of these examples include body parts or properties belonging to human beings such as a voice, a smile, etc., which are placed in **central Focus**. The mid Controller (the dative mu ‘him’ and mi ‘me’) are persons to whom ‘eyes’, ‘legs’, ‘a voice’ and ‘smile’ belong. As such, these persons might normally be in control. But in these instances, their control is diminished. The presence of se undermines the usual Control ranking. In all of them there is an outside force which affects these individuals and deprives them of their full ability to act.\(^7\)

Example (17) comes from a story in the fantasy genre. In it, the man dies but is promised another life. He reacts with repulsion when he is told that he may be reincarnated in the form of a spider. In (18), two young women visiting a big city are exhausted after a full day of sightseeing. Staying at the home of a distant relative, they barely manage to pretend to listen to the conversation during an evening meal. Fatigue and drowsiness are the super-agents that affect the girls and make their eyes close. In (19), a woman experiences an earthquake and she freezes. She probably feels like running but she is not in control of her legs. Se blocks the infer-
ence that the legs are in control of the event of ukočiti which without se is glossed in English as ‘to restrain, to hold back, to brake’. The presence of se blocks certain inferences and encourages others. It says that the control relationships among participants are atypical or unexpected. The inference is never mechanical and varies from one example to the next depending, to an extent, on the meaning of the verb and the surrounding context.

To conclude, the distribution of se cannot be explained by invoking a priori message categories. Impersonal is not a category of Serbo-Croatian and se is not consistently associated with impersonal messages. The same can be said for the categories passive and reflexive. Since se is only sometimes passive, impersonal or reflexive, we should not take any of them to define se. We can understand the distribution of se only if we observe the direct effect of the meaning central Focus on the message, and how that is different from messages in which users of the language exploit the other meanings of the complex System of Participant Focus interlocked with Number, Gender, Discourse Referent and Control.

Notes

1. Serbo-Croatian is the term commonly used in linguistic literature. All the examples cited in this article come from the Serbian literature. Se is analyzed synchronically as the clitic form of sebe; diachronically se is independent of sebe.

2. The form si, absent in Standard Serbian but found today in Standard Croatian as well as Croatian dialects and some transitional dialects of Serbian close to Macedonian and Bulgarian, has not been analyzed here. It is part of different set of grammatical systems.

3. To my knowledge, other than Boranić’s (1899) study, there has not been extensive analysis of se. Se is generally discussed in studies of pronouns, middle verbs, reflexives, impersonals and passive voice, as well as in grammars, dictionaries and encyclopedias.

4. Đorđević (1988) postulates an unusually high number of se’s. Grammars generally give between three and seven different functions of se. For a Columbia School analysis of the semantic substance of Control, see Huffman (1997).

5. In utterances in which there is only one participant on the scene, such as on spava (‘he is sleeping’) or ona trči (‘she is running’), Control is inferred and not signaled. Even though participants in Focus in the events of ‘sleeping’ and ‘running’ play rather different roles, inferring the Control role is irrelevant because there is only one participant and therefore no Control ranking is needed. However, in examples such as on čita knjigu (‘he is reading a book’), the situation is quite different. In the event of ‘reading’, the referent of ‘he’ is a high Controller and the book is signaled to be a low Controller (a patient role). In utterances with more than one participant, the Control ranking is crucial in arriving at the right message.

6. When these verbs occur without se, the nuance contributed by se is absent. For example, odreći is glossed in English as ‘to deny’ (rather than ‘to renounce, to waive’) and osvrati as ‘to turn’, etc.
7. For a discussion of the Dative of Possession, see Huffman (1997: Ch. 3).

Data Sources


References


CHAPTER 11

The difference between zero and nothing

Swahili noun class prefixes 5 and 9/10

Ellen Contini-Morava
University of Virginia

“It is primarily by its contrast to non-zero elements that an element zero is distinguished from 'nothing'”. (Haas 1957: 42)

Most Swahili noun class prefixes are overt, e.g. m-toto/wa-toto 'child/children'. However, for Classes 5 and 9/10 an overt prefix occurs only in certain morphophonemic contexts. Despite superficial similarities, only the Cl. 5 prefix should be analyzed as zero (meaningful absence); Cl. 9/10 nouns simply lack a prefix. Evidence includes differences in singular-plural patterns and in derivational productivity of prefix absence. Prefix absence indicates (inherent or derived) Cl. 5 membership and singular number. The availability of prefixless Cl. 9/10, outside the normal class and number systems, helps preserve the coherence of the noun class system, and allows zero to convey meaning in the case of Cl. 5.

1. Introduction

Ever since Saussure (1922: 124) pointed out that a linguistic system allows the possibility of zero, or significant absence, there has been controversy about how and when it is legitimate to postulate such a unit. The main problem, of course, is the fact that zero has no material substance, so proof of its existence depends entirely on analogical reasoning. The usual analogy that is invoked is similarity of pattern-role with elements that do have material substance, as shown in the epigraph to this chapter. But it is not always obvious which non-zero elements should serve as the basis for the analogy. Linguists generally agree that the best case for zero is within a closed, paradigmatic class in which all members except one show some kind of overt marking. Saussure's own example, widely repeated, is the genitive plural of the Czech word žena 'woman', which unlike all the other forms in its declensional paradigm, lacks a distinctive suffix. The problem becomes murkier, though, when
you move out of this kind of closed system and into the grey area between inflection and derivation. This is the issue I will discuss in the present chapter, using the example of the noun class prefixes of Swahili.

2. Morphological background

In Swahili, as in other Bantu languages, noun stems carry a prefix that indicates the class and number of the noun. Noun class and number are also indexed on other elements such as adjectives, demonstratives, verbs and so on. The relevant forms are listed in Table 1 below:

<table>
<thead>
<tr>
<th>Class #</th>
<th>Nominal</th>
<th>Adjectival</th>
<th>Pronominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>(traditional Bantu numbering)</td>
<td>(affixed to noun stems)</td>
<td>affixed to to adjective &amp; numeral stems</td>
<td>(affixed to V stems as subj. or obj. and to “pronominal” stems: Demonstr., Poss. &amp;c.)</td>
</tr>
<tr>
<td>1</td>
<td>m-</td>
<td>m-</td>
<td>yu-; ye-; w-; a-; m- (depends on stem)</td>
</tr>
<tr>
<td>2</td>
<td>wa-</td>
<td>wa-</td>
<td>wa-</td>
</tr>
<tr>
<td>3</td>
<td>m-</td>
<td>m-</td>
<td>u-</td>
</tr>
<tr>
<td>4</td>
<td>mi-</td>
<td>mi-</td>
<td>i-</td>
</tr>
<tr>
<td>5</td>
<td>Ø or ji-</td>
<td>Ø, j- or ji-</td>
<td>li-</td>
</tr>
<tr>
<td>6</td>
<td>ma-</td>
<td>ma-</td>
<td>ya-</td>
</tr>
<tr>
<td>7</td>
<td>ki-</td>
<td>ki-</td>
<td>ki-</td>
</tr>
<tr>
<td>8</td>
<td>vi-</td>
<td>vi-</td>
<td>vi-</td>
</tr>
<tr>
<td>9</td>
<td>Ø or N-</td>
<td>Ø or N-</td>
<td>i-</td>
</tr>
<tr>
<td>10</td>
<td>Ø or N-</td>
<td>Ø or N-</td>
<td>zi-</td>
</tr>
<tr>
<td>11/14</td>
<td>u-</td>
<td>m-</td>
<td>u-</td>
</tr>
</tbody>
</table>

Before proceeding, let me point out a few things about the data in Table 1. First, the concept of “noun class” as used in Bantu scholarship includes both a prefix on a nominal stem and a characteristic pattern of associated prefixes occurring on other elements in the sentence that index the noun. The noun classes are numbered according to the traditional Bantu numbering system whereby separate numbers are assigned to singular and plural classes. In general, odd-numbered classes are singular and the immediately following even-numbered classes are their “corresponding
Chapter 11. The difference between zero and nothing

plurals”. There are some exceptions to the general pattern: the nouns of Classes 9 and 10 are invariable in form, and for these classes number is distinguished only in the “concordial” prefixes. Class 11/14, the result of a merger between two historically distinct classes, does not have its own “corresponding plural”–plurals of nouns from this class may follow the pattern of either Cl. 6 (e.g. ugonjwa/magonjwa ‘illness/illnesses’) or Cl. 10 (e.g. ubao/mbao ‘board/boards’). This fact will become relevant below, as the plural forms of Cl. 11/14 nouns are usually cited as evidence that Class 10 has a nasal prefix. Finally, it should be noted that the Adjectival prefixes (those in the third column of Table 1) are almost all identical to the nominal prefixes (those in the second column). This point will also become relevant below.

The forms I am concerned with here are the ones in boldface on the table: the nominal prefix of Class 5, listed as zero alternating with ji-, and those of Classes 9 and 10, listed as zero alternating with a nasal. (The meaning of the upper case N will be explained below.) I will argue that the traditional analysis of the Cl. 5 nominal prefix is correct, but that Cl. 9–10 nouns, contrary to the traditional analysis, carry no prefix and are not morphologically marked. That is, Cl. 5 has a zero, whereas Cl. 9–10 have nothing. Let me start with the actual forms of the nouns in these classes. These are listed in Table 2 below:

Table 2. Formal alternations, Classes 5 and 9-10

<table>
<thead>
<tr>
<th>a.</th>
<th>Distribution of overt prefix (ji-) in Cl. 5:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ji- before monosyllabic stems, e.g. jiwe ‘stone’ (pl. mawe)</td>
</tr>
<tr>
<td></td>
<td>irregularly j- before vowels, e.g. jambo ‘thing’ (pl. mambo), but cf. ua ‘flower’, pl. maua</td>
</tr>
<tr>
<td></td>
<td>Elsewhere no overt prefix, e.g. shoka ‘axe’ (pl. mashoka)*</td>
</tr>
<tr>
<td>b.</td>
<td>Distribution of putative nasal prefix in Cl. 9 and 10:</td>
</tr>
<tr>
<td></td>
<td>stressed homorganic, syllabic nasal before monosyllabic stems, e.g. /m̩yι/ ‘grey hair[s]’;</td>
</tr>
<tr>
<td></td>
<td>/ʼn ηa/ ‘wax’;</td>
</tr>
<tr>
<td></td>
<td>homorganic prenasalization of voiced obstruents before polysyllabic stems, e.g. /m^bεɡu/ ‘seed[s]’; /dεdɛfɛ/ ‘bird[s]’; /goma/ ‘drum[s]’; /jʊdɛu/ ‘peanut[s]’; /m^vua/ ‘rain[s]’;</td>
</tr>
<tr>
<td></td>
<td>palatal nasal before vowel, e.g. /noka/ ‘snake[s]’;</td>
</tr>
<tr>
<td></td>
<td>Elsewhere no nasal, e.g. /pua/ ‘nose[s]’, /siku/ ‘day[s]’.</td>
</tr>
</tbody>
</table>

* Absence of overt prefix must be posited even for most Cl. 5 nouns with initial j-, whose plural forms add ma- rather than replacing the j, e.g. jembe ‘hoe’, pl. majembe or jogoo ‘rooster’, pl. majogoo.

† In some dialects, initial voiceless stops and affricates on Cl. 9-10 nouns are pronounced with additional aspiration, reflex of historical nasal prefix, e.g. /p^hua/ ‘nose[s]’. The nasal prefix was first devoiced, then metathesized from pre-oral stop voicelessness to post-oral stop voicelessness (Nurse & Hinnebursch 1993: 158).
To repeat: the argument I am making is that the morphophonemic patterns seen in Cl. 9–10 (those listed in b in Table 2) do not represent a separable morpheme, but at most are a set of constraints on the phonological structure of nouns in this class. These constraints are relics of historical changes that have reduced the integrity and salience of what was formerly a prefix, to the point where synchronically it has ceased to exist. In what follows I will first outline the relevant historical developments, then provide the synchronic arguments. After that I will suggest some functional reasons why the changes affected this particular noun class rather than others.

3. Historical background

Historical accounts of Swahili reconstruct a prefix *N- for Classes 9 and 10 in Proto Northeast Coast (PNEC) Bantu (see e.g. Nurse and Hinnebusch 1993: 338). This prefix assimilated in point of articulation to a following consonant. Before voiceless consonants the nasal was first devoiced, then lost before fricatives; before voiceless stops it persisted in some dialects in the form of aspiration of the now-initial stop (see second note to Table 2). Before vowels the nasal was realized as a palatal nasal, as it is in modern Swahili.2

In PNEC Bantu, nominal prefixes in certain discourse contexts were preceded by a “preprefix” or “augment”, a deictic element that was lost in Swahili. In most Bantu languages that have retained augments, these are identical or very similar in form to the pronominal prefix found on demonstratives, possessives, and so on (see Table 1)–indeed, demonstratives are thought to be the source of the augments in Bantu (see Greenberg 1978). In the case of Classes 9 and 10, the augments were differentiated for number, just as the modern pronominal prefixes are, but the Cl. 9–10 nominal prefixes were identical. Thus Nurse and Hinnebusch (1993: 338) reconstruct the augment+prefix of Cl. 9 in PNEC Bantu as *i+N-; for Cl. 10 *zi+N-.

At an earlier stage, then, number for nouns in Classes 9 and 10 could be differentiated by the augment, in discourse contexts where it was used, but this is no longer the case in modern Swahili.

Interestingly, although there has been debate among Bantu scholars as to the number and identity of the noun classes of Proto Bantu, there is widespread agreement that the nominal prefixes for Classes 9 and 10 were identical (reconstructed as *n- by Meeussen 1967: 97, *ny- by Guthrie 1971: 9).3 It is also generally assumed that Classes 9 and 10 were a singular-plural pair in Proto Bantu, even though there is considerable variation in singular-plural noun class pairings within the Bantu family (see Maho 1999: 255–7). That is, the anomaly of the Cl. 9–10 nouns being undifferentiated for number, unlike the nouns of other classes, is not just a his-
historical accident of Swahili, but rather reflects a longstanding pattern in the Bantu language family.

4. Arguments for and against a separable prefix for Cl. 9–10

The most obvious argument in support of the claim that Cl. 9–10 does not have a separable prefix in modern Swahili is the fact, already alluded to, that Cl. 9–10 nouns are invariable in form, unlike the other noun classes in which a singular prefix is replaced by a different prefix in the plural (see Table 1). As pointed out in the last section, this may well have been the case since Proto Bantu times. One reason that scholars have always posited separate, homonymous nominal prefixes for Classes 9 and 10 is based on analogy with the other noun classes, which have distinct singular and plural prefixes, as well as the fact that a number distinction is made in pronominal elements that agree with Cl. 9–10 nouns, as shown in Table 1. (Other reasons will be discussed below.) Although in Swahili the absence of overt prefix is a characteristic of most nouns in Cl. 5 as well (see Table 2a above), Cl. 5 nouns nevertheless always have a distinct plural form, e.g. *shoka/mashoka* ‘axe/axes’, whereas Cl. 9–10 nouns do not: *ndoo* ‘bucket[s]’, *pua* ‘nose[s]’. So the rationale for positing a zero prefix in a Cl. 5 noun like *shoka* ‘axe’ is that in other noun classes, such as 1, 3, and 7, a singular prefix is replaced by a different prefix in the plural. Clearly this analogy does not hold in the case of Cl. 9–10, where there is no singular-plural alternation.

A second argument against positing a prefix for Cl. 9–10 has to do with the relative lack of salience of this putative prefix, by contrast with the other Swahili noun class prefixes. The putative prefix of Cl. 9–10 is less phonologically salient than any of the other noun class prefixes: an overt nasal occurs only on a minority of Cl. 9–10 nouns, and then only in assimilated form in which the nasal coalesces with a following stop, yielding a prenasalized stop rather than a sequence syllabic nasal + imploded stop as with the other nasal prefixes (those of Classes 1 and 3).

Both arguments mentioned above suggest that learners of Swahili are less likely to perceive a distinction between prefix and stem in the case of Cl. 9–10 than in the case of the other noun class prefixes. Why, then, do Swahili grammars recognize a separable prefix for them? The arguments for positing a nasal prefix (alternating with zero) for Cl. 9–10 are listed under (A–C) below.

First, recall that Cl. 11/14 does not have its own corresponding plural class. Plurals of nouns in this class may be in either Cl. 6 or Cl. 10. The argument related to the present discussion has to do with these Cl. 10 plurals:
A. Cl. 10 plural forms of nouns whose singulars are in Cl. 11/14 (with the prefix *u*) show the typical Cl. 10 morphophonemic patterns (cf. Table 2b above) replacing the singular *u*-prefix of Cl. 11/14. For example: /udēvu/ ‘a beard hair’ (Cl. 11/14), /c̱devu/ ‘beard’ (Cl. 10)–stem with initial voiced obstruent; /ufavu/ ‘rib’ (Cl. 11/14), /m̱bavu/ ‘ribs’ (Cl. 10)–stem with initial voiced obstruent; /wimbo/ ‘song’ (Cl. 11/14, *w*- is prevocalic alternant of *u*), /ŋimbo/ ‘songs’ (Cl. 10)–vowel-initial stem; /ufi̱o/ ‘stick’ (Cl. 11/14), /fi̱o/ ‘sticks’ (Cl. 10)–stem with neither voiced obstruent nor vowel.

If we regard the plural forms here as bearing a Cl. 10 prefix, then by analogy Cl. 9 also has a prefix, because 9 and 10 are identical.

A second argument that is usually used for recognizing a Cl. 9–10 prefix has to do with adjectives. Adjective stems in Swahili carry a prefix that depends on the class of the noun they agree with, and recall that the adjectival prefixes are almost all identical to the corresponding nominal ones (see Table 1 above).

B. Adjective stems agreeing with Cl. 9–10 nouns mimic the morphophonemic patterns found with nouns of this class, e.g. -/baya/ ‘bad’: /m̱be̱g̱u m̱baya/ ‘bad seed[s]’–voiced obstruent initial stem; -/ḏo̱g̱o̱/ ‘small’: /m̱be̱g̱ u c̱do̱g̱o/ ‘small seed[s]’–voiced obstruent initial stem; -/e̱usii/ ‘black’: /m̱be̱g̱ u ŋe̱usii/ ‘black seed[s]’–vowel initial stem; -/fupi/ ‘short’: /m̱be̱g̱u fupi/ ‘short seed[s]’–stem with neither voiced obstruent nor vowel initial.

So here, too, we appear to have alternation between the Cl. 9–10 pattern and the other prefixes.

The third argument usually put forward in support of positing a nasal prefix for Cl. 9/10 nouns has to do with derived diminutive and augmentative nouns:

C. In Swahili, a noun can be made diminutive or augmentative by replacing its usual nominal prefix with that of Cl. 7 (*ki*-, diminutive) or that of Cl. 5 (*ji* ~ Ø, augmentative): for example, *mtoto* ‘child’ (Cl. 1), *kitoto* ‘little child’ (Cl. 7), Ø-*toto* ‘big, clumsy child’ (Cl. 5). If a nasal-initial Cl. 9–10 noun is made diminutive or augmentative, the initial nasal is replaced by the diminutive or augmentative prefix respectively. For example: /c̱de̱g̱e/ ‘bird[s]’, /ḵi̱de̱g̱e̱/ ‘little bird’, /ḏe̱g̱e̱/ ‘big bird’.

Once again, the claim that Cl. 9–10 nouns carry a prefix rests on the alternation between a nasal (in nouns that have one) and something else. On the face of it, these would seem to be compelling arguments in favor of a (synchronic) prefix.
Let us address each argument in turn (see Herbert 1978 for similar arguments in reference to other Bantu languages).

First, in A above it was suggested that the Cl. 10 plural forms of nouns whose singulars are in Cl. 11/14 were potential evidence for a segmentable Cl. 10 prefix. However, for many nouns of Cl. 11/14, their Cl. 10 plural forms cannot be predicted by deleting $u$- and applying Cl. 9–10 morphophonemic rules to the remainder. For example:

**Table 3.** Unpredictable Cl. 10 plurals for Cl. 11/14 singular nouns

<table>
<thead>
<tr>
<th>Cl. 11/14 singular</th>
<th>Actual Cl. 10 plural</th>
<th>Plural predicted by rules (see Table 2b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>uso ‘face’</td>
<td>ɲuso</td>
<td>*ŋso (monosyllabic stem)</td>
</tr>
<tr>
<td>[cf. uʃwe ‘string’]</td>
<td>ɲwwe</td>
<td>conforms to monosyl. stem rule</td>
</tr>
<tr>
<td>uati ‘house-pole’</td>
<td>ɲbati</td>
<td>*ŋati (vowel-initial stem)</td>
</tr>
<tr>
<td>uʃango ‘betelnut sheath’</td>
<td>bango</td>
<td>*m=bango (voiced obstruent initial)</td>
</tr>
<tr>
<td>uʃeti ‘stanza of poem’</td>
<td>bɛti</td>
<td>*m=beti (voiced obstruent initial)</td>
</tr>
<tr>
<td>uzio ‘fish trap’</td>
<td>ɲuzio</td>
<td>*m zio</td>
</tr>
</tbody>
</table>

The data in Table 3 suggest that the plural forms are not derived from the singulars by regular rule, but are rather learned as suppletive alternants of the lexical stems (using “suppletive” in the sense intended by Hudson 1974, as quoted by Herbert 1978). This weakens the case for regarding the Cl. 10 plural form as having a separable prefix.

Let us move now to the argument stated in B, about the forms of adjectives that agree with nouns of Cl. 9–10. As pointed out in B, adjectives agreeing with nouns of Cl. 9/10 mimic the phonological patterns found in the nouns of this class. If an adjective stem satisfies the conditions that call for an initial nasal segment in a Cl. 9/10 noun (i.e. if the adjective stem begins with a voiced obstruent or vowel), the Cl. 9/10 adjective will have an analogous initial nasal segment. If those conditions are not met (i.e. if the adjective stem begins with a voiceless obstruent), no initial nasal is found. However the argument that the forms of Cl. 9/10 adjectives provide evidence for a nasal prefix on Cl. 9/10 nouns only holds if one regards the set of prefixes occurring on adjectival stems to be identical to those occurring on nominal stems. But there are also compelling arguments for regarding the adjectival prefixes as distinct from the nominal prefixes. These are listed in D below.

D. i. Not all adjectival prefixes are identical to nominal prefixes. As shown in Table 1 above, the adjectival prefix for Class 11/14 is $m$-, whereas the nominal prefix for that class is $u$-.
ii. Adjectives agreeing with Cl. 5 nouns show a more regular pattern of prefix distribution than nouns of that class do (see Table 2a above): all vowel-initial adjective stems agreeing with a Cl. 5 noun have initial j- (e.g. j-eusi ‘black [Cl. 5]; j-embamba ‘thin [Cl. 5]’), whereas vowel-initial Cl. 5 noun stems only irregularly have j- (cf. Table 2a and first table note).

iii. Adjectival stems occur freely with the prefixes of all noun classes, supplying learners of Swahili with multiple forms distinguished by the various prefixes (unlike noun stems, which show alternation only in classes that distinguish singular and plural). Therefore speakers could recognize and segment a nasal prefix for adjectives that index a Cl. 9–10 noun (when a nasal occurs) without necessarily doing the same for the noun stems. Speakers could also simply memorize the Class 9–10 adjectival forms rather than creating them by means of a rule. This would not put a great burden on the memory, since the lexical class “adjective” is very small in Swahili, as it is in Niger-Congo languages generally (cf. Welmers 1973, Ch. 9; Nurse & Hinnebusch 1993: 305).

If the adjectival prefixes are treated as distinct from the nominal prefixes, there is no reason to extrapolate a Cl. 9–10 nominal prefix on the basis of the behavior of adjectives.

Finally, a rebuttal to the third argument, having to do with derived diminutive and augmentative nouns. In C it was stated that if a Cl. 9/10 noun has an initial nasalized segment, the nasal is replaced by the diminutive or augmentative prefix in derived nouns, thus potentially providing evidence for a segmentable nasal prefix in the noun. However, just as the plural forms of Cl. 11 nouns are not entirely predictable based on Cl. 9/10 phonological patterns (see Table 3), so also the replacement of a Class 9–10 initial nasal in derived diminutive and augmentative forms is not completely predictable. Examples:

<table>
<thead>
<tr>
<th>Cl. 9-10 noun</th>
<th>Derived (Cl. 5) augmentative</th>
<th>Derived (Cl. 6) augm. pl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>joka 'snake'</td>
<td>joka 'large snake'</td>
<td>majoka 'large snakes'</td>
</tr>
<tr>
<td>naŋna 'tomato'</td>
<td>naŋna ~ jiŋna 'large tomato'</td>
<td>maŋna 'large tomatoes'</td>
</tr>
<tr>
<td>undo 'hammer'</td>
<td>undo ~ jundo ~ jiundo 'lg. hammer'</td>
<td>undo ~ majundo ~ majiundo 'lg. hammers'</td>
</tr>
</tbody>
</table>

* My thanks to Charles Bwenge for collecting data on augmentative forms of Cl. 9–10 stems among his fellow Swahili instructors at the African Language Teachers Association workshop in 2003.
The source nouns all have an initial palatal nasal, found when the following segment is a vowel (cf. Table 2, b above). The first example conforms to the rule that replaces the initial nasal with the Cl. 5 prevocalic prefix j- in the singular. However, in the other two cases the nasal is sometimes retained in the augmentative instead of being replaced. Here, too, the alternation between nasal prefix and something else, where present, may be regarded as a case of suppletive stem alternants rather than as productive replacement of one prefix by another. Thus none of the arguments listed under (A-C) conclusively proves that Cl. 9–10 nouns should be treated as a sequence of prefix + stem.

5. Further arguments against a separable prefix for Cl. 9/10

In the last section I considered several arguments that have been put forward in support of the claim that there is a Cl. 9–10 prefix, even though all nouns of this class are invariable for number. I argued that none of these arguments leads inexorably to this conclusion. In this section I will bring up two more arguments against a separable prefix for Cl. 9–10. The first has to do with loanwords. For all other noun classes, loanwords add the respective class prefix (or the initial syllable is treated as a prefix and replaced with the appropriate plural prefix). By contrast, loanwords in Classes 9–10 never add a nasal even if they satisfy the conditions for doing so (see Table 2, b above). Examples:

Table 5. Examples of loanwords in classes other than 9-10

<table>
<thead>
<tr>
<th>Loanword</th>
<th>Source:</th>
</tr>
</thead>
<tbody>
<tr>
<td>mafaci 'a German' (Cl. 1)</td>
<td>deutsch (German)</td>
</tr>
<tr>
<td>mibobo 'cashew tree' (Cl. 3)</td>
<td>bibo (Portuguese)</td>
</tr>
<tr>
<td>khatari 'tin oil lamp' (Cl. 7)</td>
<td>batil 'lamp wick' (Persian)</td>
</tr>
<tr>
<td>usukani 'rudder' (Cl. 11/14)</td>
<td>sukān (Arabic)</td>
</tr>
</tbody>
</table>

Compare loanwords in Classes 9–10:

Table 6. Examples of loanwords in Classes 9-10

<table>
<thead>
<tr>
<th>Loanword (invariable form):</th>
<th>Form predicted by rules:</th>
<th>Source:</th>
</tr>
</thead>
<tbody>
<tr>
<td>baraka 'blessing[s]'</td>
<td>*baraka</td>
<td>bārak (Arabic)</td>
</tr>
<tr>
<td>dola 'dollar[s]'</td>
<td>*dola</td>
<td>dollar (English)</td>
</tr>
<tr>
<td>adāfu 'politeness'</td>
<td>*adāfu</td>
<td>adab (Arabic)</td>
</tr>
</tbody>
</table>
Presumably loanwords are exempt from the phonotactic constraints that constitute the historical remains of the former Cl. 9–10 prefix because these remnants no longer have the synchronic value of a morpheme.

A final argument against a separable prefix for Cl. 9–10 has to do with derivational use of noun class prefixes. All Swahili noun class prefixes, with the sole exception of Cl. 9–10, can be used productively to create nouns from nominal, adjectival or verbal stems. Examples:

Table 7. Examples of noun class prefixes used to derive nouns

<table>
<thead>
<tr>
<th>Derived noun:</th>
<th>Source:</th>
</tr>
</thead>
<tbody>
<tr>
<td>m-linda ‘guard [N]’ (Cl. 1)</td>
<td>-linda ‘guard [V]’</td>
</tr>
<tr>
<td>m-kutano ‘meeting’ (Cl. 3)</td>
<td>-kutana ‘meet [V]’</td>
</tr>
<tr>
<td>θ-toto ‘large child’ (augmentative—Cl. 5)</td>
<td>m-toto ‘child [Cl. 1]’</td>
</tr>
<tr>
<td>ki-toto ‘small child’ (diminutive—Cl. 7)</td>
<td>m-toto ‘child [Cl. 1]’</td>
</tr>
<tr>
<td>u-zuri ‘beauty’ (Cl. 11/14)</td>
<td>-zuri ‘beautiful (Adj.)’</td>
</tr>
</tbody>
</table>

Note especially the use of zero, or deletion of noun class prefix, as a productive way of deriving augmentative nouns, which are treated syntactically as nouns of Cl. 5. The example shown in Table 7 is θ-toto ‘large child’, derived from the Cl. 1 noun mtoto ‘child’, with the prefix m- deleted. Although the putative nasal prefix of Cl. 9–10 is also supposed to have a zero alternant, the absence of a prefix where one might otherwise be expected is always interpreted as an instance of a Cl. 5 augmentative, never of a derived Cl. 9–10 noun.

6. Conclusion

In conclusion, the arguments presented above strongly suggest that Cl. 9–10, alone among all the noun classes, should be regarded as lacking a prefix in contemporary Swahili. By contrast, Cl. 5 does have a prefix, whose most frequent alternant is zero. This apparently anomalous situation can be understood if one takes various functional factors into consideration.

First, Cl. 9–10 are the most semantically heterogeneous of all the Swahili noun classes (Contini-Morava 1994, 1997). Both for this reason and because of the low salience of the Cl. 9–10 prefix due to its erosion over time, this class has also been the most hospitable to loanwords whose initial segments do not fit the phonological patterns of recognizable class prefixes (Nurse & Hinnebusch 1993: 351–355). The influx of loanwords into Cl. 9–10 in turn has helped weaken whatever semantic and phonological coherence it originally had. Since Cl. 9–10 is also unique in
failing to differentiate singular from plural, it is outside the normal noun class system in this respect as well.

A second, related point has to do with the pronominal concords of Cl. 9–10. The semantic neutrality of this class has been extended to its pronominal concords also. The singular concord i- (see Table 1) is used in antecedentless contexts such as impersonals (inaonekana ‘it seems’) and to index expressions that are not normally nouns but are temporarily being treated as such (alijibu ‘hodi’ yake ‘he answered his “hello”’–Contini-Morava 1996). In pidginized varieties of Swahili, where the agreement system has been lost, the Cl. 9 forms become generalized to noun modifiers (adjectives, demonstratives, possessives, etc.) that would otherwise vary according to class (Heine 1973). That is, Cl. 9–10 has come to serve as a residual class, a home for nouns that do not fit readily into other classes and an agreement pattern for entities that are not prototypical nouns.

In conclusion, a zero, or significant absence, can be most easily recognized (and therefore can reliably convey its meaning) within a closed set of oppositions in which all other alternatives are overt marks of some kind (Saussure 1922: 124; cf. Sanders’ 1988 “overt analogue criterion”). The existence of the prefixless Cl. 9–10, outside the normal noun class and number system of Swahili, helps preserve the coherence of the noun class system as a whole. More specifically, it makes possible the use of zero for the unambiguous derivation of augmentative nouns, which as pointed out earlier are placed in Cl. 5. Exactly because Cl. 9–10 has nothing, Cl. 5 can have a zero.

Notes

1. Swahili forms will be cited in Standard Swahili orthography, except where the phonological details are relevant to the argument, in which case they will be cited in phonemic transcription.

2. Apparently the palatal is a reflex of a high front vowel or palatal glide that followed the initial nasal at some time prior to PNEC and was lost before consonants: reconstructions of Proto-Bantu sometimes include this element (e.g. Guthrie 1971: 9, who posits *ny- as the prefix for Classes 9 and 10).

3. An exception is Cole (1967), who argues that Proto Bantu Class 10 actually had a “compos-ite prefix” *li-ni-, in which a plural prefix *li- was superimposed on the singular prefix *ni- of Cl. 9. Cole thus agrees with Meinhof (1948), who reconstructed the prefix of Cl. 9 as *ni- and that of Cl. 10 as *li-ni, but did not explain the nature of *li-. Cole acknowledges, however, that “in very many Bantu languages (probably the majority), there is no difference in overt form between nouns of Cl. 9 with prefix */ni-/ and those of Cl. 10–9 with prefix */li-ni/, i.e., the singular and plural forms are phonologically identical” (1967: 126).
4. Retention of initial $j$-following the plural prefix $ma$- is the norm for $j$-initial augmentatives. It is also common for $ma$- to be added to, rather than replacing, initial $j$- before Cl. 5 noun stems (see first note to Table 2).

References


A semantic analysis of the Swahili suffix *li*

Robert A. Leonard and Wendy Saliba
Hofstra University

This Columbia School analysis of the Swahili suffix *li* rests on the distinction between meaning and message, as the proposed hypothesis does not simply categorize message types as previous analyses have done, but rather posits a single meaning which accounts for the various messages to which *li* contributes. Our hypothesis, unlike others’, accounts for all instances of *li*, as well as for what has been called “double-*li*”. The analysis posits a new type of Control meaning. Specifically, *li* instructs the hearer to interpret a non-high controller at a higher, more potent level of control than if *li* were not used.

1. A new analysis of *li*

This paper presents an analysis of the Swahili postverbal particle *li*, one of a number of forms traditionally analyzed as extending or modifying the meaning of a verbal stem. Our analysis differs from earlier ones in that we identify a meaning for *li* by examining the contribution of *li* to the overall communicative content of the utterance.

We take the analytical approach of the Columbia School: that language’s communicative function determines its structure. The basic resources of language, and the objects of analysis, are signals paired with meanings. Each meaning “can have a wide variety of exploitations in the task of communicating a wide variety of messages, the variation itself always depending on other factors in the context, not on the meanings alone” (Diver 1982: 21). This distinction between meaning and message will prove pivotal in the analysis below.

In this paper, we will begin in Section 2 by reviewing previous analyses of *li*, and by explaining how our analysis differs. Then, in Sections 3 and 4 we will present our hypothesis, which will be explicated and supported with naturally occurring data both with and without *li*. In Section 5, we will consider what has been
called the “double-li construction”, and we will show how the double-li phenomenon supports our analysis. Sections 6 and 7 will address, respectively, the issues of usage categories and the distinction between meaning and message that is crucial to our analysis.

2. Generative and traditional treatments enumerate uses of li

Previous accounts of li generally describe not the form’s meaning but the various English translations of its uses. Li has previously been analyzed as changing the meaning of the verb to a meaning in which the verb is applied towards someone. It has been called (1) “prepositional”, e.g., by Ashton (1944), because it is often translated into English with a preposition; (2) “applied”, e.g., by Polome (1967); and (3) “dative”, by Vitale (1981).

Of previous analyses, only Port (1981) breaks with this. Like the present analysis, Port hypothesizes that li has a single meaning, but he claims a different meaning from ours, and his treatment of reduplicated li differs as well. Port’s meaning is “add a role to the lexical verb that is less active than the agent.” This we found inadequate, for, as explained below, li does not merely add a role, but rather raises the participancy of some non-high controller on the scene. Further, the main thrust of Port’s paper is that productive reduplicated li does not exist and that most of what analysts thought were reduplications are actually frozen lexical stems. This we will also discuss below, and, we believe, refute.

Vitale (1981), a generative transformational treatment, considers li an overt marker of the dative (44 ff.), although he notes that li with some kinds of verbs (verbs of motion followed by a motivating goal) are difficult to formalize and “their relation to true datives is semantically more tenuous than has hitherto been assumed” (60 fn.). Vitale does not aim to demonstrate that li is a dative marker; rather, the analysis starts with that assumption and provides examples to illustrate that presumed fact. Further, Vitale does not analyze data such as (b) below, where there is no overt indirect object in the gloss.

Traditional approaches to li enumerate and categorize uses of the form. Ashton (1944), the standard Swahili reference and didactic grammar, offers the following list of message categories she terms uses. In this list, Ashton gives citation forms of verbs, e.g., imba (in bold), and examples of each verb with li (in its morphophonemic variants), e.g., imbia, (also in bold):

(a) To express “to do to, for, or on behalf of someone, or to the detriment of such a one”…

imba ‘sing’ Watoto waliuimbia nyimbo
Chapter 12. A semantic analysis of the Swahili suffix *li*

The children sang songs to us...

*lima* ‘cultivate’  
*Amemlimia shamba*  
He has cultivated his plantation for him...

*haribu* ‘destroy’  
*Atatuharibia furaha*  
He will spoil (for us) our pleasure.

(b)  
To express motion *towards*…

*panda* ‘climb’  
*Walipopandia ile*  
When they climbed up the baobab trees...

*geuka* ‘turn’  
*Yule mfalme akali- geukia Jua akasema*  
The king turned to the Sun and said...

(c)  
To express *purpose*…

*jenga* ‘build’  
*Kamba hizi kama ni za kujengea, hazitafaa.*  
If these ropes are for building with, they will be useless.

*kata* ‘cut’  
*Nataka kisu cha kukatia nyama.*  
I want a knife for cutting meat.

(d)  
To express *finality or completeness*…

*tupa* ‘throw away’  
*Mikebe hi niitupe? (no li)*  
Am I to throw away these tin cans?  
*Itupie mbali. (with li)*  
Throw them right away.

Ashton (1944: 218–220)

Ashton paints a good picture of *li*’s semantic richness and the problems of rendering into English:

In these illustrations verbs have been chosen which allow of a fairly precise translation, but it is often impossible to express the meaning conveyed by the root and the suffix *[li]* in a single word which suits all contexts... Moreover, as Swahili words frequently require a much wider vocabulary in English, the reader must on no account restrict the meaning of words to those given in the vocabulary or in a particular context. Many words have a figurative meaning also. (Ashton 1944: 216)

Such cataloging (e.g., purpose, finality, completeness, beneficiary, maleficiary, motion towards) results in a non-exhaustive categorization of fragments of the wide range of messages to which the meaning of *li* can contribute.
3. Our hypothesis

3.1 Meanings of the Control System

We propose that *li* signals a meaning in the semantic substance of *Control*. Huffman explains that grammatical meanings of Control indicate:

> to what degree the participant ... contributes to bringing about the event. Participants contributing to higher degree may actively control the event, or, if they do not actually initiate the event, at least make a major contribution to setting it up. Those contributing to a lesser degree will be responsible for the precipitation of the event in a less direct way; this may consist of a subtle form of motivation, or even range down to almost no contribution at all. (Huffman 1985: 80; see also Huffman 1997)

We hypothesize that Swahili *li* is the speaker’s instruction that the hearer should interpret a non-high Controller at an enhanced level of Control. That is, we hypothesize the hearer is instructed to:

- a. analyze the communicative situation;
- b. identify which entity is High Controller;
- c. identify which other entities exercise some lower degree of Control (non-high Controllers), and;
- d. choose one whose raising would contribute to a coherent, plausible message.

Thus *li* instructs the hearer to **interpret a non-high controller at a higher, more potent level of control** than if *li* were not used. This meaning, which we will refer to as **elevate non-high controller**, might at first glance seem a bit subtle, since it does not correspond to any well-known, well-worn structuralist meaning sanctioned by tradition or familiarity.

**Non-high** Control is a relatively wide area of meaning. It is defined only in relation to **high** Control, and can include many roles that in absolute terms might be described as patient, beneficiary, motivator, etc., depending on context. Notice further that while *li* raises a non-high Controller, it does not raise an entity out of non-high into high. As illustrated in Figure 1, *li* only raises within non-high Control.
Figure 1. Effect of *li*

Thus – and this is important because it separates our hypothesis from most previous analyses – *li* does not signal an *absolute* level of control. Rather, *li* moves an entity up on a sliding scale of Control, and the hearer infers what would be a plausible message in a given context.

3.2 An illustration: ‘pick up and carry’ vs. ‘deal with’

The effect of *li* may be quite striking in translation. For instance, in the following examples, the verb *chuku* translates differently depending on whether *li* is present. In (1) *chuku* (without *li*) is rendered in English as ‘(physically) carry, take’, whereas in (2), the same verb, with *li*, is translated as ‘deal with’:

(1) Baba, labda mkimchukua pole pole atasema
   Father perhaps you-ki-her-carry slow slow she-FUT-speak
   ‘Father, perhaps if you carry her gently she will talk.’
(2) Baba, labda mkimchukulia pole pole atasema (PM 13)
   Father perhaps you-ki-her-carry-*li* slow slow she-FUT-speak
   ‘Father, perhaps if you deal with her gently she will talk.’

Note that in (1), the non-*HIGH* controller (‘her’) is considered baggage, while in (2), with *li*, that controller, elevated, is more actively participating.

Table 1 shows the analysis of examples (1) and (2), with regard to the *HIGH* Control entity (‘you’), the non-*HIGH* Control entity (‘her’), and the presence or absence of *li*, as well as the English gloss.

<table>
<thead>
<tr>
<th>Example</th>
<th>HIGH Control (Subject Marker)</th>
<th>Non-HIGH Control (Object Marker)</th>
<th><em>li</em>?</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>‘you’</td>
<td>‘her’</td>
<td>no</td>
<td>‘you carry her’</td>
</tr>
<tr>
<td>(2)</td>
<td>‘you’</td>
<td>‘her’</td>
<td>yes</td>
<td>‘you deal with her’</td>
</tr>
</tbody>
</table>
The difference in the English glosses might give the impression that what is being changed from (1) to (2) is the action named by the verb. But this is just an artifact of the English translation; what we have in Swahili is the same verb with and without li.

To summarize, examples (1) and (2) form a minimal pair. In (1), the non-high Controller is interpreted merely as a load to be carried, a patient being acted upon, but in (2), with li, she is a participant involved in the give-and-take of being dealt with, someone whose opinion must be taken into consideration. In (2) the woman is more in control than the dead weight of (1). This difference in level of Control must be attributable to the presence of li, since the sentences are otherwise identical.

4. Li is a signal of a unitary meaning

As we have said, our analysis presents li as the signal of a single meaning, elevate non-high controller, and posits that this meaning contributes to a range of messages depending on context. The following are four examples, (3)–(6), of the verb acha ‘leave, let be’ in different contexts, taken from modern plays, in which we see a continuum of messages of ever higher control on the part of the non-high Control participants.

In (3) the non-high participant (‘me’) has been left – abandoned, really. This is the low end of the range of roles that can be performed by non-high participants. In this passage, a young woman from a tiny village finds herself pregnant and alone in the big city. She says to herself, “the time has come that I no longer have any desire to live. What is the use of living if all that I get is bitterness and a breaking of my heart.” Then she cries out:

(3) Mama kwa nini ukaniacha nikutane na haya yote? (PM 7)
Mother for what you-ka-me-leave I-meet-SBJ with this all

‘Mama, how could you leave me to meet up with all this?’

The young woman’s role here is that of patient, a victim abandoned to evil circumstance with no say in or control over the leaving.

Similar but not identical to (3) is (4), in which the same morphology is used (no li). The person (‘me’) in the non-high role in (4) is of slightly higher Control than the speaker in (3) in that she is the one making the suggestion that the leaving be done, but her suggestion is highly diffident. Here, a very young woman is thrown out of her house by her mother, and a male friend marries her out of pity. Soon after, she walks in on her new husband holding another woman. She isn’t angry, only sad and resigned, as she knows that they never really had a marriage. He had only done her a favor. She says to him:
Chapter 12. A semantic analysis of the Swahili suffix *li*

(4) Naona ... mimi naona (pole pole) ... ingalikua bora ukaniacha (EH 39)
   I-PRES-feel I-I-PRES-feel slow slow it-COND-be better you-ka-me-leave
   ‘I feel ... I feel (spoken slowly) ... it might be better if you were to leave me.’

Were there a *li* on the ‘leave’ verb in (4) (*ukaniacha* would be *ukaniachia*), it would signal that the wife was playing a raised Control role, was more involved in the ‘leaving’. Even though it is she who is publicly acknowledging their mutual realization that the time has come to end their marriage, it is the husband’s place to leave her. A raising of her Control with *li* would be interpreted as less diffidence on her part.

Contrast examples (3) and (4) with (5), in which the non-high Controller very actively prompts the leaving and loudly beseeches the high Controller to grant permission for it. This attempt to precipitate the leaving warrants a *li*. In (5), there has been a pointless, bloody fight between two clans. Everyone is injured, but one young man wants to re-enter the fray. His grandfather (the clan elder) forcibly sits him on the ground and pins him down. The grandfather says, “Don’t heap your stupidity on me right now. You have vexed me and thrown away respect for my age and advice. I am fighting for your life – they will kill you, by God!” The young man replies:

(5) Hawawezi kuniua. Niachie mimi nikawaue. (PM 33)
   NEG-they-able-NEG INF-me-kill me-leave-li me I-ka-them-kill-subjunctive
   ‘They’re not able to kill me. Leave me be and I’ll go kill them.’

The non-high Controller in (5), where we find *li*, strongly urges his grandfather to leave him and so participates more in the ‘leaving’ than do his counterparts in (3) or (4).

In (6), the non-high Controller is in an even higher Control position than in (5): in this instance the same young man not only wants the leaving, but is actively struggling to free himself. This is the high end of the non-high Control range and is marked by a double instance of *li*. This passage, from the same play as (2) and (3), is just prior to the events in (5). The men return from a fight, and the young man wants to fight more. The stage directions say that he is being held back by his grandfather, and that he “tries to leap from the hands” that are restraining him “to go back and fight”. He says:

(6) Babu niachilie. Niachilie nikawafundishe si ... sikubali. (PM 33)
   Grandfather me-leave-li-li me-leave-li-li 1-ka-them-teach-SBJ I ...I NEG-agree
   ‘Grandfather, leave me be. Leave me be, I’ll go and teach them, I... I won’t stand for it.’
In (6) the young man is actively trying to leap from the arms of his restrainer. Thus, he contributes even more actively to wanting to be left alone than in (5). In (5) the verb has one li; in (6) there are two instances of li.

Table 2 shows the communicative result of li with respect to the act of leaving. Examples (3) and (4) do not contain li; example (5) does; and example (6) includes two instances of the form. While the two instances of li are referred to in the literature as the “double-li construction”, we see here that the presence of double-li suggests an increase even higher than a single li in the level of Control of the non-high participant.

<table>
<thead>
<tr>
<th>Example Number</th>
<th>HIGH Controller</th>
<th>Non-HIGH Controller</th>
<th>li?</th>
<th>Role of the non-high participant in the ‘leaving’</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3)</td>
<td>‘you’</td>
<td>‘her’</td>
<td>no</td>
<td>abandoned</td>
</tr>
<tr>
<td>(4)</td>
<td>‘you’</td>
<td>‘her’</td>
<td>no</td>
<td>meekly suggests</td>
</tr>
<tr>
<td>(5)</td>
<td>‘(you) imperative’</td>
<td>‘me’</td>
<td>yes</td>
<td>actively wants</td>
</tr>
<tr>
<td>(6)</td>
<td>‘(you) imperative’</td>
<td>‘me’</td>
<td>yes-double</td>
<td>actively struggling</td>
</tr>
</tbody>
</table>

5. **Double-li supports the Control hypothesis**

We analyze double-li, as in (6), as two instances of li with additive effect; that is, two occurrences of li raise the non-high Controller more than one li. It is interesting that although double-li is formally two instances of li, previous analysts have not viewed double-li as two instances of the meaning of li. For example, Polome (1967: 85) says the “reduplicated form” stresses “completion” of the action identified by the verb with a single unreduplicated li. It is evidence in favor of the present analysis that double-li has demonstrably the same effect on the message as what we have observed for single li, though to a greater degree.

Other analysts have treated double-li as a construction with a meaning independent of that of single li. For example, Ashton (1944) says – rather confusingly – that double-li is the same morphology as single li, but with a different meaning: the “double-prepositional form” is the “Same in form as certain ... [p]repositional suffixes” (p. 214) but not otherwise connected to the prepositional, having a meaning of “continuance of the action and persistence in carrying it out” (p. 244). Ashton shows some understanding of the message, here persistence, but little of how the message was arrived at. That is, she does not see that there is a connection among all the messages (that they can be explained by positing a single meaning
for *li*), or that these examples contain a repetition of the same meaning, with additive result. Example (6), for instance, shows the persistence Ashton mentions – think of how frenzied and persistent the activity level of the non-HIGH controller had to get before we saw two *li* forms. But the message is persistence not because two *li* forms in a row constitute a construction, but because two *li* forms in a row raise the non-HIGH controller higher than does one *li*.

Port (1981), in an otherwise excellent and insightful article, seems not to allow that two *li*’s may occur with a single verb:

Many descriptions of the applied form in Swahili discuss cases of apparent reduplication of the suffix ... with the suggestion that they are somehow emphatic versions of the simple suffix (Polome 1967: 85).... I propose that they actually represent a frozen lexical suffix... this double suffix is not generally productive and has only a handful of examples – maybe only 10 or 15. These verbs, too, can add another [li] that acts like the normally productive one, [and if a verb already] contained a productive [li] then we could not add another…. (1981: 74–75)

None of Port’s reduplicated examples are in context, so it is difficult to assess his judgments. He does not make clear what the “10 or 15” examples are.

A major thrust of Port’s article is that many of the analytical problems concerning *li* stem from the confusion of productive *li* with “frozen lexical stems that look the same or similar”; that is, Port says that there are “many lexical relics that resemble [li] but are related to it only historically” (1981: 71). Port gives the example (p. 74) of *kimbia* ‘run’, that looks like an instance of a verb *kimba* plus *li*, but *kimba* does not occur and therefore, apparently by definition *kimbia* cannot consist of a verb plus *li*. Thus *kimbilia*, which looks like a double-*li* and would be analyzed as double-*li* if *kimba* occurred, actually contains only one instance of *li*.

But all such examples have to be examined carefully. While there may indeed be frozen, non-productive stems that at first blush look like productive double-*li*, one need only compare naturally occurring examples like (5) and (6) – almost a minimal pair on this account – to establish that there exist instances of productive double-*li*.

6. Many usage categories are but manifestations of the single *li*

As we have seen, the meaning that *li* contributes is always the same, though the form has different effects on the overall message in different contexts – both linguistic and real-world. This difference in overall message between the non-*li* and the *li* examples in the data we have seen is not easily formulated in traditional terms, and, as we have noted, has caused analysts to account for *li* by offering lists of uses.
Earlier, we discussed Ashton’s (1944) list of uses of *li*, which she presents as four totally distinct categories:

a. To express “to do to, for, or on behalf of someone, or to the detriment of such a one.”

... 

b. To express motion towards...

... 

c. To express purpose...

... 

d. To express finality or completeness. When thus used the verb is generally followed by an adverb such as *sana* ‘much’, *mbali* ‘far away’. (Ashton 1944: 218–220)

Let us review each of these uses, and explain how each is actually a manifestation of the single meaning of *li*. We will augment Ashton’s examples with additional data. Ashton’s first use (a) is the one classically referred to as the beneficiary or maleficiary use. The maleficiary is illustrated in the following pair of examples:

(7) Bwana Se ... Bwana Sembuli alinikosea kwa nguvu na sasa mimi ni mja mzito. (PM 14)

Bwana Se...Bwana Sembuli he-PAST-me-wrong-*li* by force and now I am person heavy

‘Bwana Se ... Bwana Sembuli did wrong to me by force and now I am pregnant.’

The morphology of this example suggests that the speaker believes she motivated the action of the verb as an unwilling source of desire, and thus *li* is deployed by the writer. By contrast, example (8), from the same play, contains the same Swahili verb, *kosa*, without *li*. The presence and absence of *li* results in different English translations of *kosa* – in (7) ‘do (someone) wrong’ and in (8) ‘miss’.

(8) Tumewakosa labda wamefuata njia nyingine, wako wapi? (PM 30)

We-PERF-them-miss maybe they-PERF-follow road another they where

‘We’ve missed them, maybe they took another road – where are they?’

In (7) and (8), we find the same verb, *kosa*, whose core meaning covers ‘lack, miss, mistake, be wrong, do (someone) wrong.’ It is clear from the context in (8) that the group being missed is not avoiding the others; they have no role in the missing other than that they are, by chance, somewhere else. The non-high Controller (‘them’) thus does not contribute as much to the ‘missing’ as the speaker (‘me’) in (7), who,
by existing as a woman in the man’s house and perceiving herself as an object of unwanted desire, views herself as the likely cause of the fact that she was ‘wronged’.

A motivated wronging, directed at the motivator, is the classic maleficiary message. This message arises not because *li* means maleficiary but because, given the right context and lexical items, the maleficiary message is an appropriate inference from *li*’s meaning of instructing the hearer to *interpret a non-high controller at a higher, more potent level of control*. Here, the more potent level of control is to motivate the wronging. Thus, the maleficiary role is one possible categorization of the message to which *li* contributes.

Two other uses of *li* described by Ashton (1944), (b) and (c) above, are “to express motion towards…” (p. 219) and “to express purpose…” (p. 220). These uses are exemplified by the stage directions in (9):

(9) Anamwona Sembuli anamwenda kutaka kumpiga. (PM 19)
   he-PRES-him-sees Sembuli he-PRES-him-go-*li* INF-want INF-him-hit
   ‘He sees Sembuli, he goes for him wanting to hit him.’

In this example, the messages “motion towards” and “purpose” are both in evidence. The targeted Sembuli, who is also a maleficiary, motivates the going. Here, we can clearly see the connection between the message effects of creating “motion towards” and creating “purpose”, and the meaning we have posited for *li* (*interpret a non-high controller at a higher, more potent level of control*).

With a higher Control role, the message effect is that the participant is a motivator, and serves as a purpose for an action.

Ashton’s last use, (d) above, is “to express finality or completeness”. Ashton (1944: 220) offers the following examples with *tupa* ‘throw away’, one without *li* and one with *li*:

(10) a. Mikebe hii niitupe?
    cans these I-them-throw
    ‘Am I to throw away these tin cans?’

b. Itupie mbali.
    them throw *li* far
    ‘Throw them right away.’

Ashton gives us no context for these examples, and without context we cannot find a reason for *li* to be there. We might, however, expect to find such a phrase as *Itupie mbali* in a context in which the cans possess some quality of being noxious, useless, or disgusting that would motivate the ‘throwing’ immediately, that is, motivating their instant and absolute expulsion, as opposed to mere normal removal of garbage.
The opposite emotion, love, fuels *li* in the following, in which the raised Control level resulting from *li* also fits Ashton’s (1944: 220) message category of “finality”. Her example was “throwing something right away.” In (11), also from the play *Hatia*, a character says they must ‘leave aside’ something that actively demands attention. In this example, a teenage girl from a small village, alone in the big city of Dar es Salaam, learns she is pregnant and searches the streets for Juma, her lover. She finds him, but he seems annoyed. “Why would you be looking for me?” he asks. She replies, “Ah Juma, my love. You are asking me why I look for you,” and then she says:

(11) Haya yote tuyaachili mbali. (PM 8)
this all we-it-leave-*li*-SBJ far
‘All this we must leave aside. I have big news to tell you, Juma.’

‘All this’ that must be ‘left aside’ is the entity whose Control level is raised by *li*. ‘All this’ is her great love, which must be ‘left aside’, because she has the important news of her pregnancy to tell Juma. The love quite actively demands attention, and as such is marked as heightened Control, a more active participant in its being put aside.

In the following example, *li* contributes to a kind of message that is not included in any analyst’s list of message categories, nor does it seem to be explainable by the frameworks of earlier hypotheses. Example (12) is part of the introduction to the play *Wakati Ukuta*, and gives the time and place setting of the play’s events:

(12) MAHALI: Mambo haya yalitokea katika mkoa ya pwani. (EH 8)
PLACE: matters these they-PAST-issue-*li* province of coast
‘PLACE: These things happened in Coast Province.’

The play’s title translates as “Time is a Wall,” which means, as a main character explains, if you try to fight Time, it will be like hitting a wall – it will be you who is hurt. The overarching theme is the collision between changing times and tradition. The place, Coast Province itself, strongly motivates the events that occur in the play, for it is in Coast Province where live the strictest, most orthodox, most traditional Muslim Swahilis. Without *li*, the very common verb *toka* gives a reading of an ‘issuing forth’ or ‘leaving’ that neither affects nor is affected by what it issues into. In (12) *li* signifies that the ‘issuing forth’ of these matters involves a very motivating participant: its venue, the traditional Coast Province. The example communicates ‘These things happened in Coast Province (as one who knows the place might expect).’ While this example does not fit into Ashton’s (1944) list of uses seen above, it is explainable by the meaning hypothesis we have proposed.
7. Meaning vs. message

We have seen that Ashton's (1944) well-known grammar of Swahili appears to analyze li by examining English translations and discovering the many messages that can be communicated with this form. Failing to see any more unifying principle than that li “gives a prepositional concept to the simple form of the verb,” (p. 218) Ashton gives a list of disparate and extremely general uses (pp. 217–221), which attempt to categorize fragments of the wide range of messages. Ashton's list bears a striking resemblance to the inventory of uses traditionally assigned the Latin dative.

Similarly, Vitale's (1981) generative treatment of li examines the messages associated with this form, and since some of them coincide with the a priori notion of dative, he assigns that label even though the dative function fails to account for certain data. However, even at the lowest level of description, lists are of limited use: some examples fit into several different usage headings at once and there will be others, like (12), not covered by the list of uses.

Traditional terms like prepositional and dative are actually cover terms for a list of disparate uses. The methodology of traditional analysis thus is predisposed to create disunified lists, and each analysis unifies only in that it calls all these uses by a single name.

Nevertheless, traditional analysts do not view such list-making as problematic. A major reason for this, as noted before, is that traditional, including generative, grammars do not make a crucial distinction: between meaning and message. As Huffman explains in his analysis of the French clitic pronouns:

An important distinction is made between meaning – an actual resource of the language, and message – the total content which the meanings help the hearer to infer. This is how humans manage to use such a limited array of linguistic resources to communicate a much larger number of messages. In our analysis, we united such diverse notions as beneficiary and maleficiary not by trying to abstract over them, but by recognizing that the contribution of lui-, mid Degree of Control, is the same in both cases; and that the inference of a message of ‘maleficiary’ as opposed to ‘beneficiary’ comes not from lui, but from other elements of the context and from other knowledge the speaker has. (Huffman 1985: 514)

Thus our analysis also “unite[s] such diverse notions as beneficiary and maleficiary” and many other even more diverse, “not by trying to abstract over them,” (Huffman 1985: 514) but by realizing they are ad hoc categories of the limitless number of different messages that can be inferred from the signaling of a single meaning (ELEVATE NON-HIGH CONTROLLER) in different contexts. This all-important difference between meaning and message distinguishes Columbia School analyses. Our meaning ELEVATE NON-HIGH CONTROLLER may be characterized,
to use Huffman’s words (1995: 204, 205), as “sparse”, and “imprecise,” but it is emphatically not an abstraction across uses. To abstract across uses limits the analysis to a predetermined set of uses that typically derive from either groupings of English translations of messages (e.g., Ashton 1944), or inherited *a priori* Latinate categories (e.g., Vitale 1981). Other exploitations of the communicative functions of forms become, like example (12), invisible, as there is an infinity of different possible messages, depending on context (see Leonard 1985, 1995). Even Ashton, an expert abstracter elsewhere, does not suggest any abstract meaning that could apply to her own data of *single* instances of *li*, let alone double-*li*. This is less surprising when we realize that Ashton is aware of a good part of the range of *li*’s possible messages, but simply cannot find a single abstraction, *because* she is working *from* the messages.

Abstraction across uses goes in the wrong direction: the essential difference between Columbia School and traditional and generative treatments is that the latter two

> make their goal the analysis of the message *after* it has been understood rather than *how* it is that the message *gets* understood... Only the message is accessible to a traditional analysis and thus its categorization of a form’s semantic content is almost necessarily multiple. (Leonard 1980: 210)

In conclusion, previous analyses cannot account for all the data we have seen regarding *li*. The analysis we have presented here is successful because it does not begin with an analysis of message categories. Rather, the examples adduced above show that *li* has a single, invariant meaning, *elevate non-high controller*, that contributes to the wide variety of message types that we have seen.

**Notes**

1. Our examples, except for the invented (1), are from two modern Swahili plays. English translations are ours. The following abbreviations are used:

   - FUT Future
   - PRES Present
   - NEG Negative
   - PERF Perfect
   - SBJ Subjunctive
   - COND Conditional
   - INF Infinitive

2. *Ki* is a tense unique to Swahili and other Bantu languages and not easily glossed. *Ki* is usually translated as ‘if’ in the non-past, ‘-ing’ in the past.

3. While the underlying form is most likely /li/ and is realized as such in the present example, phonological processes cause /l/ to drop in most environments. The vowel is realized as either /e/ or /i/ depending on whether the vowel(s) of the verb stem is (are) mid height or not: if the other vowels are mid height, the form is /e/; if not, /i/.
4. *Ka,* like *ki,* is a tense unique to Swahili and other Bantu languages and not easily glossed. *Ka* is widely misanalyzed as meaning ‘sequential action’ (it is often called the ‘*and then*’ tense) but that is just one of the messages inferred from its meaning in a probability system, as is the "go and" message in *ka* plus subjunctive, e.g., *tukatafute maji* ‘let us go [but never ‘come’] and look for water’, and its probability message, as in *bora ukaniacha* ‘[it would be] better if you were to leave me’ in example (4) above. See Leonard (1980).

Data Sources


References


The structure of the Japanese inferential system

A functional analysis of *daroo*, *rashii*, *soo-da*, and *yooda*

Hidemi Sugi Riggs
Soka University of America

This paper provides an analysis of Japanese inferential auxiliaries (*daroo*, *rashii*, *soo-da*, *yooda*) that enables us to explain all occurrences of these words in discourse. Despite extensive research on these forms, grammatical analysis has not yet successfully explained their distribution. This paper demonstrates that the prevailing view of these forms as *evidentials* is insufficient. Instead, we postulate a new hypothesis in which the speaker’s choice of auxiliary is based on his or her desire either to strongly present inferential information or to draw attention away from it. In this analysis, the language user is seen as more active in the interpretation of messages than is traditionally assumed.

1. Introduction

In modern Japanese there are four forms traditionally analyzed as *inferential auxiliaries*, namely *daroo*, *rashii*, *soo-da*, and *yooda*. These forms appear to have similar uses, and their translations into English frequently overlap. Respectively, the auxiliaries are commonly translated into English as ‘I think, probably’ (*daroo*); ‘seem’ (*rashii*); ‘I have a feeling that’ (*soo-da*); and ‘seem, look like, seemingly’ (*yooda*). Researchers generally distinguish the subtle differences among the forms in terms of *evidentiality*. That is, the type of evidence and the degree of the speaker’s certainty are the primary method of differentiating among the words’ meanings.

This paper argues that this prevailing view of these forms as *evidentials* is unsupported. Instead, we will offer a new analysis in which the meaning of an individual auxiliary form is defined according to its relationship with the other...
auxiliary forms, and in which the language user is seen as more active than in traditional grammatical analysis. The hypothesis is based on an analysis of data collected from approximately two thousand occurrences of the forms. The sources of these tokens are seventeen books of essays, six novels, five works of nonfiction, weather forecasts on two different television stations, e-mail messages, and online chat room transcripts.

2. Previous studies and problems

2.1 Previous research

Despite extensive research into these forms, grammatical analysis has not successfully explained their distribution. Researchers often present the auxiliaries as able to be used interchangeably in certain contexts (e.g., Sakurai 1972; Kashioka 1980; Johnson 1994). In introducing “epistemic markers” including *daroo*, Ohta (1991: 213) states, “Givón (1982) describes evidentials as markers showing the speaker's evaluation of the truth-value of a proposition and revealing the speaker's placement of the proposition in epistemic space.”

Givón (1982: 42), one of the best-known analyses of these forms, provides a clear taxonomy demonstrating three degrees of speakers’ certainty, namely, Lowest, Medium, and Highest. Lowest Certainty reflects an irrealis state in which the speaker is presenting a hypothetical statement and for this reason cannot employ an evidential form in expressing his thoughts. For this reason, Givón refers to the Lowest Certainty category as “evidentiality impossible.” Medium Certainty reflects a realis state in which the speaker presents an inference for which he has evidence that supports it. Statements reflecting this category require the use of an inferential auxiliary, and for this reason, Givón terms this “evidentiality required.” Finally, Highest Certainty includes deictically obvious statements, presuppositions, revelations, and “generic knowledge shared within the culture” and coded in the language. As with the Lowest Certainty category, Highest Certainty expressions do not require the use of evidential forms or constructions, according to Givón. Therefore, evidential forms are required only when speakers are presenting inferences in which they have only a moderate degree of certainty and for this reason require a certain degree of evidence to strengthen their propositions.

Contrary to Givón’s classification, *rashii* is chosen by speakers to convey both an eyewitnessed event and a questionable event, as we will see in examples (1) and (2) below:
Chapter 13. The structure of the Japanese inferential system 241

(1) Kinoo obaachan no tokoro ni ikimashita. De, hochoki da kedo, jitsu wa moo motteiru-rashii. (Misete- morattara, kasutamu meido no mono des-hita.) (email message)

Yesterday grandma GEN place LOC go.PAST As for hearing aid COP COJ fact TOP already possess-rashii (shown.given.then custom made GEN thing) COP.PAST

‘I went to Grandma’s house yesterday. As for the hearing aids that we discussed [though we assumed otherwise] in reality she already owns [rashii] a pair. (She showed it to me, and it was a custom-made thing.)’

(2) Sono sheikuya-san de das-areru sheiku tte jitsu wa amerikan shooto heyaa kara tsuku-tte-iru rashii nya (online resource)

The shake shop-hor LOC serve-PSV shake QT fact TOP American short hair from make.and.be rashii EX

‘To tell the truth, the shake-stand seems [rashii] to serve shakes that are made from American shorthair cat!’

The first use of rashii in (1) contradicts the conventional account of the word because the speaker actually witnessed the event. On the other hand, the use of rashii in (2) appears to convey highly unlikely information, though the context also supports Givón’s (1982) category of “evidentiality impossible,” in which low certainty expressions are defined as incompatible with the use of evidentials. Contrary to Givón’s classification, rashii is chosen by the speakers to convey both an eyewitnessed event (1) and a reverie (2) with the revelatory word jitsu wa ‘in fact’.2

According to the prevailing view, soo-da is used when inferences are drawn from the speaker’s gut feeling, anticipation of a future event, conjecture of another person’s state of emotion or physical state, and so forth: ureshikute shini soo-da ‘I feel like I will die of supreme bliss’. However, soo-da is regularly used to express hypothetical and counterfactual propositions, such as natsu ni pan-suto nannka haitei tara, moe-agari-soo-ni naru ‘If I wore pantyhose in summer, I would almost burst into flame.

According to Kindaichi (1953), the auxiliaries that have inflection (yooda, rashii, and soo-da) indicate that the speaker’s conjecture is based on an objective stance such as visual or hearsay evidence. On the other hand, the auxiliary that has no inflection (daroo) indicates that the speaker’s conjecture is based on a subjective stance such as a guess.3 Kashioka (1980) hypothesizes that a speaker uses yooda when conjectures are based on direct observation of one’s own intuitive judgment. Note that Kashioka’s interpretation of yooda does not agree with Kindaichi’s, because intuition cannot be construed as an objective source of information.
Teramura (1984) posits a single core (prototypical) meaning for each form. Daroo’s core meaning is posited to be ‘I think that,’ indicating that the speaker’s conjecture is not based on observation. Soo-da’s core meaning is posited to be ‘I feel that,’ indicating that the speaker’s conjecture is based on his gut feelings on a given matter. Teramura presents the difference between yooda and rashii as the degree of subjectivity.

Aoki (1986) states that yooda is used when the speaker has some visible, tangible, or audible evidence collected through his own senses, enabling him to make an inference, while rashii is used when the evidence is circumstantial or gathered through sources other than one’s own senses. The analytical base here is, once again, the type of information source.

Makino and Tsutsui (1994) state that rashii usually expresses the speaker’s conjecture based on what the speaker has heard or read (i.e., secondhand information), while yooda expresses the speaker’s conjecture based on firsthand, reliable information; that is, yooda involves the speaker’s reasoning process. Makino and Tsutsui, however, contradict themselves when they state that rashii can refer to evidence that is seen. Soo-da expresses the speaker’s conjecture based on visual information, and it can only be used when the speaker directly observes something.

Maynard (1990) says rashii conveys the likelihood that a particular conjecture is true, without necessarily committing the speaker to believing the certainty of it, partly because the speaker’s speculation is based on secondhand information. In other words, when using rashii, unlike yooda, the speaker distances himself from his statement, conveying that he is not fully committed to or totally responsible for it. Daroo conveys doubt in the speaker’s guess.

Shirota (1998) defines a clear-cut dichotomy between two groups. According to Shirota, yooda and rashii indicate that the speaker’s inference is based on some particular information, while soo-da and daroo indicate that the speaker’s conjecture is not based on any particular information. Although at first glance this classification of yooda and rashii shows a close resemblance with the hypothesis to be proposed here, Shirota relies on the notion of “speaker’s territory” vs. “hearer’s territory”: he says that yooda is used when the speaker’s inference is based on information within his territory, while rashii is used when the speaker’s inference is based on information beyond his territory. However, defining “territory” can be problematic. In (3), the speaker is making his inference in the interlocutor’s presence:

(3) “Um, moji wa onna no yooda na.” (Gonin taiho: Mistaken Arrest)

Um letter TOP woman GEN yooda CP

‘Um, it looks like [yooda] a woman’s writing.’
Two detectives are examining a libelous, anonymous letter found against a possible suspect. The detectives are attempting to find out what kind of person wrote it. The speaker is making an inferential judgment based on concrete visual evidence in the presence of another detective. How can the speaker claim that the situation belongs to his territory? In addition to this pragmatic constraint, the appearance of the final particle *na* apparently indicates that the speaker is soliciting agreement from the interlocutor who is also looking at the letter. The data of the present research suggest that the territory hypothesis does not account for the grammatical distinction between *yooda* and *rashii*.

Despite the preponderance of accounts based on evidentiality (i.e., firsthand versus secondhand information), there are various instances found in the data that show the inadequacy of that approach.

### 2.2 Further counter-examples to the evidentiality approach

#### 2.2.1 *Rashii and yooda*

There is much that is problematic for those who advocate a firsthand versus secondhand opposition between *yooda* (e.g., “intuitive” or “eyewitness”) and *rashii* (e.g., “hearsay”). Examples (4) and (5) cannot be explained by the prevailing approaches based on evidentiality:

(4) Shooji ni hi ga atat-teru. Yarai no ame ga agatte, dooyara kyo wa tenki ga ii rashii. (*Iro to tsuya no nihon bunka: ‘Colors and a Gloss of Japanese Culture’*)

*shoji* screen *LOC* sunlight *SBJ* hit-*be* since last night *GEN* rain *SBJ*stop. and likely today *TOP* weather *SBJ* good *rashii*

‘The sunlight shines on the *shoji* screen. The rain that fell last night has gone, and it *seems* [rashii] that it is a beautiful day today.’

In this excerpt, the speaker makes his inferential judgment based on his own experience recognizing the bright *shoji* screen hit by sunlight.


about victim *LNK* identification *SBJ* settle came that stage *P* Fukuoka-preference-police *DAT* minutely investigation *ACC* requested the result *SBJ* now report *PST* CMP *COP* victim *TOP* Tamura Wakako *DAT* almost mistake-NGT *yooda*
‘Their primary investigation into the victim’s identification is finished. At this stage, they asked the Fukuoka police department for a more detailed investigation. Now, they have just received the report. “It seems [yooda] to be almost one hundred percent positive that the victim was Tamura Wakako, as we expected.” ’

Despite the fact that the speaker received secondhand information (i.e., the police report), he nonetheless uses yooda. This further weakens the traditional claim that yooda is reserved for firsthand information.

2.2.2 Daroo and soo-da
According to conventional grammar, propositions presented with daroo are interpreted as the speaker’s irresponsible guess or a guess that is clouded by the speaker’s doubt. However, the data in this paper show that daroo is often used to indicate speakers’ conviction toward their conjectural judgments. Example (6) is one of these instances:

(6) Kadoo to iu mono o watashi wa shiranai. Kadoo ni tsuite no hon o yonda koto mo nai. Shikashi aru yo, osoku made shigoto o shiteita toki, shiin to shizumarieta nakade kasuka da ga nanika ga kuzureru yoona oto ga shita... Shodana ni oita kabin no bara no hana ga kuzureta no dearu. Watashi wa sono toki, kore ga nihonjin dokutoku no kadoo no honsitsu daroo to omotta. (Haru no hana: Spring Flowers)

Kadoo (flower arrangement) QT say thing ACC I TOP know NEG flower arrangement DAT about GEN book ACC read PST matter also NEG But one night late until work ACC do.be.PST time ONM QT be.silent in subtle be but something SBJ collapse like sound SBJ do.PST bookcase LOC put. PST vase GEN rose GEN flower SBJ collapse PST NOM be I TOP that time this SBJ Japanese peculiar GEN flower arrangement GEN essence daroo QT realize PST

‘I don’t know anything about Kadoo, nor have I ever read a single book about it. One night I was working on my writing and all of a sudden in the complete silence I heard a very subtle sound as if something collapsed.... It was the sound of the petals, which were in the vase on the bookcase, suddenly falling from the rose. At that moment I realized [daroo] that that is the very essence of the Japanese art of flower arrangement, which is peculiar to Japanese.’

This excerpt is found in a book of essays. The writer confesses that he is ignorant of the Japanese art of flower arrangement. Nonetheless, he concludes the essence of the art is the transient nature of the flowers’ mortality. His conjecture is derived
from his gut feeling. Although the writer does not have direct factual evidence with which to support his view, his certainty of this proposition is quite high.

Likewise, example (7) is an anomalous use of *soo-da* from the perspective of the previous analyses based on evidentiality because the speaker is not making an inference but rather verbalizing something in his imagination. The instance will demonstrate that even though the thought is based on the appearance of the referent, it is by no means visual evidence or eyewitness conjecture.

(7) Kotoshi wa tenko fujun de shimai kanete-ita ga, konya nado wa sekiyu-su-toobu mo sasugani shozainasa-*soo-da*. (*Kotoba shunju*: Words in Seasons)

‘I haven’t put the kerosene stove away yet due to the unstable weather this winter. The stove looks [*soo-da*] practically “bored” on such a warm night as tonight.’

The speaker in this excerpt is describing what he felt when he saw the kerosene stove on a warm night in spring. It is not a simple conjecture based on the appearance of an object, but it is his imagination about the appearance of the kerosene stove. The kerosene stove is not the visual evidence on which the speaker’s conjecture is based. Instead, the speaker is expressing his impression of the kerosene stove, while being aware that a non-living object does not have emotions and cannot feel bored.

3. A sign-based hypothesis

3.1 Theoretical preliminaries

In Section 3.2 we present a more successful sign-based hypothesis that assigns to the language user a more active role than is usually the case in conventional accounts. This analysis does not rely on the type of information source on which the speaker draws his conclusions, but rather on the speaker’s own intentions in his presentation of the information.

Our hypothesis, like the Columbia School framework on which it is based, views language as a communicative device, rather than as a representational device. Linguistic meanings are hypothesized to be merely *rough hints* to messages, rather than a complete fraction or building block of a message. In this theory, semantics (at least the semantics of grammatical elements) is relatively sparse and abstract, so that pragmatic factors play a much larger role than they do in other approaches.5
3.2 A sign-based hypothesis

The signal-meaning analysis proposed here consists of two sets of semantic oppositions, and is graphically represented in Figure 1. The first opposition is determined by the mode of conjecture, namely, inference established versus inference non-established. The term inference established implies that there is particular information on which the speaker’s inference is based, for instance, information that a speaker previously saw or heard. However, though the speaker’s inference is based on specific information, this does not necessarily imply that the interpretation of the information is correct; the speaker may be mistaken. Inference non-established refers to an inference that has no relevant or established information. The speaker does not make his conjecture based on any specific information, but rather on his own subjective thought or opinion such as intuition, general knowledge of the world, or common sense.

Each of these two categories has subcategories that consist of two sets of oppositions, high focus and low focus. Thus, there are four different subcategories in terms of inference type and degree of speaker emphasis. In the inference established node, if a speaker makes an inference that is based on established information, then he has the option of either one of the inferential auxiliaries, i.e., rashii or yooda. It is the speaker’s task to choose the form based on the degree of emphasis he wishes to place on the inference. If he wishes to forcefully present his inference, then he uses rashii. On the other hand, if he wishes to present his inference less assertively, then he uses yooda.

Meanwhile, if the speaker makes an inference without relying on any relevant information source (i.e., inference non-established), then he may choose either daroo or soo-da, depending on how much he wishes to put emphasis on his conjecture. If the speaker wishes to present his conjecture forcefully, he may use daroo. By contrast, if he does not wish or need to put emphasis on his conjecture, he may choose soo-da. In other words, he uses soo-da when he provides unimportant information.

Figure 1 shows the one-to-one relationships between the linguistic forms and their invariant meanings. Rashii and yooda share the semantic domain inference established, but they are opposed to each other in that they signal different degrees of focus. Likewise, daroo and soo-da share the semantic domain inference non-established. The former signals high focus, while the latter signals low focus.
Figure 1. Interlock of the Japanese systems of Inference and Focus

Rather than assert that the four Japanese inferential auxiliaries have meanings that overlap with each other, the present paper accepts the Saussurean principle that different meanings systematically contrast with one another. Furthermore, the semantic categories of a language do not simply reflect the objective structure of an external reality but rather reflect the exigencies of communication. For instance, the speaker who said, *hara ga hette shini-soo-da* ‘I feel as if I will die of hunger,’ does not seriously think he will literally die of starvation, but nonetheless uses the expression to complain about his hunger. To accomplish this he needs to draw the hearer’s attention away from *shinu* ‘to die;’ thus *low focus soo-da*. This meaning is opposed to that of *daroo* as discussed in (6), above, where the falling petals occasion a profound ‘realization.’ A speaker or writer is not simply matching words to predefined referents. Rather, a speaker or writer is distinguishing unique concepts that characterize each of the four lexemes. In the words of Reid (1991):

Linguistic categories are conceptual tools for structuring and interpreting messages about an experiential reality that by itself is open to multiple characterizations. The analytical consequence of this position is that semantic analysis must focus on the communicative function of words, not their referential function. (Reid 1991: 54)
4. Testing the hypothesis

4.1 Data analysis

The data analyzed here are comprised of approximately two thousand instances of rashii, yooda, daroo, and soo-da. Table 1 sorts these data into several categories.

Table 1. Token frequency of the auxiliaries

<table>
<thead>
<tr>
<th>Category</th>
<th>Co-occurrence with</th>
<th>rashii</th>
<th>yooda</th>
<th>daroo</th>
<th>soo-da</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>moshi ‘if’</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>Weather report on TV</td>
<td>0</td>
<td>0</td>
<td>8*</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>jitsu wa ‘in reality’*</td>
<td>213</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>ki ga suru ‘I have a feeling that’</td>
<td>13</td>
<td>766</td>
<td>33</td>
<td>1827</td>
</tr>
<tr>
<td>5</td>
<td>Metaphor/hyperbole/personification</td>
<td>0</td>
<td>405</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>6</td>
<td>kara ‘because’</td>
<td>29</td>
<td>1</td>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>Other contexts</td>
<td>348</td>
<td>314</td>
<td>622</td>
<td>173</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>603</td>
<td>1491</td>
<td>788</td>
<td>2048</td>
</tr>
</tbody>
</table>

* The auxiliary actually appearing in the weather report is deshoo ‘will be’, which is widely acknowledged to be the polite form of daroo. Male speakers in casual conversation commonly use daroo. Otherwise, it is also used in written materials, such as newspapers, essays, and academic articles.

** The present research does not count the co-occurrence of jitsu wa ‘to tell the truth’ with the hearsay soo-da because while the inferential auxiliary directly attaches to the stem form of a declinable constituent, the hearsay auxiliary attaches to a completed sentence which forms the content of the hearsay information.

Category 1 is the co-occurrence with the conditional conjunction moshi ‘if’. The number represents the token frequency of the word in the second part of the conditional sentence. Category 1 illustrates a clear line between the two groups according to Type of Inference, namely rashii and yooda on the one hand, and daroo and soo-da on the other. This result supports the proposed analysis: rashii and yooda are used in contexts in which speakers make their inferences based on established information. To the contrary, the occurrence of moshi ‘if’ in the case of daroo and soo-da suggests that speakers can make an inference relying on no established information.

Likewise, in Category 2, representing data from five minutes of a television weather report, neither rashii nor yooda is found. No matter how much data meteorologists may collect, their conclusions are not based on established, concrete information.
In the case of Category 3, rashii and yooda are used in conjunction with the revelatory phrase jitsu wa, glossed in English as ‘to tell the truth,’ ‘in reality.’ Table 1 reveals a striking asymmetry between the uses of rashii and yooda with this phrase. Because of its meaning of HIGH FOCUS, rashii is favored in this context, since both jitsu wa and rashii point to the importance of the information expressed.

Category 4 is the co-occurrence of the forms with evasive expressions such as ki ga suru/kanji ga suru/omoeru ‘I have a feeling.’ All four forms are found in Category 4, but LOW FOCUS forms predominate. Contrast this LOW FOCUS on subjective feeling with the HIGH FOCUS on reason in Category 6, discussed below.

As for Category 5, no instance of metaphorical HIGH FOCUS rashii or daroo is found. According to the proposed analysis, these auxiliaries indicate that the speaker wants to present his inferences in an emphatic manner. Contrarily, metaphors are not straightforward but are compatible with hedging. For example, example (8) is a cliché which depicts a painful sensation caused by cold water. The expression indicates the speaker is cognizant of the implausibility of the implication that the coldness of water would sever a hand:

(8) Kono mizu tsumetakute te ga kire-soo-da.⁶

This water cold.because hand subj cut-soo-da op

‘This water is so cold that it is as if it would [soo-da] cut off my hand.’

Category 6 reveals not only the inadequacy of the analysis based on firsthand versus secondhand source information types, but also suggests that there are two distinctive groups in terms of their frequency. First, as the data show, all four auxiliaries occur in the clause following the conjunction kara ‘because.’ (The number for this category represents the token frequency of the word in the second part of the sentence, following kara.) This is also true of Category 4, as we have seen, but with a reverse skewing. Although daroo has been considered a mere guess marker (e.g., Makino and Tsutsui 1994), the frequent occurrence of daroo with the reasoning conjunction clearly indicates that speakers/writers use it when they make an inferential judgment through a cause-and-effect thinking process. Second, the allocations of the high frequency auxiliaries, rashii and daroo, versus the low frequency auxiliaries yooda and soo-da, match well to the HIGH FOCUS and LOW FOCUS meanings signaled by them. Rashii suggests that the speaker presents inferences based on established information. It is widely acknowledged that rashii often precedes the evidence on which the speaker’s inference is based in the same sentence or in adjacent utterances. Again, despite the common assumption that only yooda involves the speaker’s reasoning process, the data in Category 6 show that the most frequent auxiliaries used in the reasoning sentences are HIGH FOCUS auxiliaries, rashii and daroo. In contrast, these HIGH FOCUS auxiliaries are not
used in a figurative sense such as metaphor, the personification of an inanimate referent, etc., and are less often used, too, for impressionistic ‘feeling.’ We ascribe this result to an incompatibility of Focus on Type of Inference with the nature of metaphor. That is to say, as a metaphor compels us to split our awareness into two things almost simultaneously (cf. Lucas 1955 [1956]), the HIGH FOCUS auxiliaries are rationally contradictory to the expression.

4.2 Inference established: rashii vs. yooda

Building on the proposed hypothesis that speakers’ choices of using either rashii or yooda can be explained in terms of presenting (HIGH FOCUS) or reserving (LOW FOCUS) inferential judgment, we must come to a clear understanding of the circumstances in which speakers wish to explicitly or equivocally present their inferential judgments.

Let us examine some additional uses of rashii and yooda from the collected data:

(9) “Naani, chottoshita shujutsu desu. Taiji ga shindeiru rashii kara ne. Shitai o sooha surya iiwakedesu. (Honjitsu kyushin: No Consultation Today)

INTJ small operation COP fetus SBJ died-be rashii because FP Dead body ACC clean up do.if fine.COP

‘Well, it’s a minor operation. Because the fetus seems [rashii] to be dead. It will be alright if I clean up the womb.’ (Translated by Seidensticker)

In this excerpt from a novel, an obstetrician suggests a course of action to his patient. In the preceding text, the obstetrician had found that the fetus carried by his patient had died. Despite the fact that he is certain of the fetus’s fate on the basis of his own extensive observations and clinical expertise, he presents his findings using rashii. This example also pokes holes in the claim that rashii conveys a low degree of commitment on the part of the speaker. Just the opposite is true in this example, in which the doctor is highly committed to his opinion that the fetus is in fact dead. The firsthand/secondhand dichotomy cannot account for the use of rashii in this context either, because the doctor’s judgment is clearly based on his firsthand experience, namely, the whole course of the medical examination of his patient.

The proposed new analysis, which states that speakers use rashii to explicitly present unexpected information, accounts for the auxiliary’s appearance in this example. Based on the medical interview with his patient, he judges that his finding (that the fetus seems to be dead) will be new and unexpected information to his patient, so he uses rashii: INFERENC ESTABLISHED, HIGH FOCUS.
In contrast, the speaker in (10) does not present his inference as new and unexpected information to his interlocutor:

(10) “Kaze hiit-ta no ka?” Erika no hitai ni te o atete, “Netsu wa nai yooda kedo” (Kiri no mukogawa: The Other Side of the Fog)

Cold caught LNK Q Erika GEN forehead LOC hand ACC put fever TOP none yooda though

“Have you caught a cold?” He put his hand on Erika’s forehead and said, “You don’t seem [yooda] feverish…”

In this example, the speaker touches his niece’s forehead and judges that she is not feverish. This inferential judgment is based on his tactile sensation and represented by yooda. The use of yooda in this kind of situation is grammatically obligatory (cf. McGloin 1989) because the information source is the hearer herself. Although McGloin does not provide the reason why yooda is obligatory and rashii only conveys a hearsay sense, her description is pragmatically accurate. Under the given circumstance, it is extremely unnatural to consider the referent’s bodily sensation as shared knowledge. Instead, only the referent is entitled to claim ownership. According to our analysis, the use of HIGH FOCUS rashii in this context (namely, the speaker making an inferential judgment that refers to how the addressee feels) would be infelicitous, because no matter how much firsthand information (e.g., tactile evidence) the speaker possesses, he cannot contest the referent’s feeling, that is, whether or not she feels ill. Due to this pragmatic constraint, the speaker uses yooda, LOW FOCUS. In this respect, the situation in (10) differs from that in (9) where the speaker, the doctor, could forcefully present a judgment whereas the addressee, his patient, could not.

Examples (11) through (13) will provide further evidence that different degrees of speakers’ emphasis on the presentation of inferences lead to the use of different auxiliaries. The excerpts cited are from a book of essays whose target audience consists primarily of businessmen. The author is a male professor.

(11) Katei no shufu wa amari hon o yoma-nai rashii. Souiu chosa kekka ga aru… Okusangata wa hon o yondeiru no de wa nai-ka to yosoushiteita dakeni, kono suuji wa yaya igai deatta. (Kotoba shunju: Words in Seasons)

Home GEN housewife TOP no.much book ACC read—not rashii Such survey result SBJ be … Other’s wives TOP book ACC reading CMP COP P NEG-Q QT anticipated since this numbers TOP a little unexpected was

‘Housewives seem [rashii] to seldom read books. There are statistics that indicate this…. Since I had presumed that housewives spend more time reading, the statistical number referring to them is a little bit unexpected.’
In this example, a classic case of *rashii* being used to draw attention to revelatory information, the speaker points out his surprise at housewives’ poor reading habits. The speaker and referent are from very different social groups, the former being an intellectual and the latter spending much time in domestic environments. In other words, the speaker did not know how housewives spend their time, and is surprised to learn that they do not read very much. He augments this revelatory information by using *rashii* to draw focus to his findings about this unfamiliar group.

Contrast this use of *rashii* with example (12), in which the same writer invokes *yooda* to make a not-so-surprising observation about his own social group:

(12) Kangaete miru to (kangaete-mi-naku temo hakkiri shiteiru ga) wareware wa sukoshi kohii o nomi sugiru *yooda*. (*Kotoba shunju*: Words in Seasons)

Thinking try when (thinking-try-not even if clearly is but) we top a little. bit coffee ACC drink exceed *yooda*

‘When I think it over – it’s obvious even if I don’t think about it too seriously – we seem [yooda] to drink too much coffee.’

In this excerpt, the professor is musing over the amount of coffee that he and others in his social group consume. His use of the first plural pronoun *wareware* indicates that the speaker conceives himself and his audience (i.e., his readers) to be in the same social group. Because the author includes himself in the social group to which he is referring, he does not consider that his observation will be in any way surprising to his audience. For this reason, he opts to use *yooda*, since there is no revelatory information he is seeking to highlight.

In (13), the writer discusses the manners and customs of barefootedness in Japanese culture. He introduces two particular occasions in which it is customary to be barefoot, and then mentions the tradition of removing one’s footwear before committing suicide by drowning oneself. This information is “additive” (Halliday and Hasan 1976: 234); that is, it is introduced with the connective *sorekara* ‘in addition to it’ and for this reason needs no emphasis:

(13) Watakushitachi nihonjin wa gishikitekina tokoro de wa, hadashi ni naru no ga tadashii to omotteimasu. Tatami no ue de wa hadashi ni nari masu shi, ohyakudo-mairi mo hadashi ga hontoodesu. Sorekara jibunde jibun no inochi o tatoo to suru toki, tatoeba fune kara umi ni tobikomu toiu toki, nihonjin wa taitei geta nari kutsu o soko e nuide mizu ni hairu *yoodesu*. (*Kotoba no saho*: Good Manner of Words)

We Japanese top ritual places LOC TOP barefoot P become LNKSBJ correct P are believing tatami mat LNK above COR TOP barefoot be and hundred times-prayer also barefoot SBJ right And then by oneself own life ACC ex-
terminate time for instance boat ABL sea DAT jump into as time Japanese TOP usually sandal or shoes ACC there P take off water DAT enter yooda

‘We Japanese take it for granted [yooda] that in ritualistic places going barefoot is the correct etiquette. We go barefoot on tatami mats, and it is also the correct manner in which to make a wish by walking back and forth a hundred times before a shrine offering a prayer each time. And then when a person is about to drown himself, for instance when a person is about to throw himself into water from a boat, Japanese usually take off their footwear and leave them there before they throw themselves into the water.’

Yooda appears in (13) because the information presented is common knowledge to Japanese readers and will be of no surprise to them. In addition, the repetitive use of nihonjin includes the readers together in a group that shares similar knowledge. Since this particular custom is a long-term tradition practiced in Japan, the writer who was born and grew up in the society cannot present it as a newsworthy inference.

The following example (14) similarly illustrates the different degree of focus on inferences in terms of unexpected versus expected information related to out-of-group or in-group references.

(14) Watashi wa Hokuriku to-n-ne-ru ga mada tsukurikake no toki ni sono Kooji genba o mita koto ga aru…. Genba no oyakata no yoona hito wa toomawashi-ni kojiba ni onna nanka ireru mono dewa nai, to iu yoona koto o itta ga, watashi wa kikoe-nai furi o shiteita. (omission) Genba ni onna ga hairu to engi ga warui to iu meishin ga aru rashii ga, (omission) ookina jiko mo naku-te kanseishi-ta yooda. [Mazu bisho: Smile First]

‘I had an opportunity to visit the construction site of Hokuriku Tunnel when it was still under construction…. The man who looked like the boss at the site said that they should not let a woman step in the site, but I pretended that I didn’t hear it. (omission) It seems [rashii] that there is a jinx that if a woman gets in the construction site, it brings bad luck to the construction…; in reality, the tunnel seemed [yooda] to be accomplished without any significant incident.’
In this excerpt, the speaker, a female essayist, is mentioning her two inferences about the ‘jinx’ that requires that women be kept from construction sites and the fact of the successful completion of the Hokuriku Tunnel. The former is called a ‘jinx,’ in which a traditional practice of discrimination against women still exists. In the preceding context, the author mentions that she had been officially approved to go inside the tunnel, and she was offended when she heard the foreman at the construction site indicate that he did not want her to tour the tunnel because of her gender. Hence, in the first sentence she sarcastically presents the absurd jinx by using the **high focus** auxiliary, *rashii*. On the contrary, she uses the **low focus** auxiliary *yoooda* to refer to the matter-of-fact event that the tunnel was completed without significant incident. The proposed analysis posits that *yoooda*, unlike *rashii*, is used in contexts in which the speaker does not or cannot present new or unexpected information, as we saw in (13), where the speaker has no need to draw attention to information that is already shared.

In (15) the writer introduces information about duties of the emperors. These duties are first introduced with *rashii* and then with *yoooda*. Despite the fact that all information is based on the same historical documents, the author of the history book uses two different auxiliaries:

(15) Jissai no shuukyooteki na gyooji toshite kono koro no tennoo wa ine no
minori no yutakani naru koto ya hideri ni ame no furu koto nado o kami ni
inoru yakume o motteita *rashii* Niinamematsuri o okonau koto mo
tennoo no shigoto deatta *yoooda*. *(Kodai kokka no seiritsu: Establishment
of the Earliest State)*

Practical LNK religious ritual as this time LNK emperor TOP rice plants
LNK prosperous become matter P drought P rain NOM fall matter etc. ACC
god DAT pray role ACC had *rashii* Harvest festival ACC conduct matter also
emperor’s job was *yoooda*

‘As for religious rituals in this period, it seems (*rashii*) that the emperors
performed such duties as praying for the bountiful harvest of rice grain
and for rainfall during a drought. Conducting the celebration of new rice
crops also seems [*yoooda*] to have been their duty.’

Knowledge-based hypotheses cannot account for the writer’s uses of the forms
in this excerpt since the source of the writer’s information is the same in both
instances, though different forms are used. By contrast, the proposed theory can
easily account for these data. In presenting information of the emperors’
generally forgotten religious and ceremonial role in archaic society, the writer first uses
*rashii* to emphasize and draw attention to information that is certainly new to his
readers. Once this emphasis has been established, the author no longer needs to
draw attention to the consequent information, so he employs the **low focus** auxiliary *yooda* in the latter part of the list (cf. Chafe 1994: 122).

Now let’s turn our attention to another reasonable motive for using the **inference established: low focus** auxiliary (*yooda*), i.e., speaker’s intention of diminishing emphasis on his inference because it is uncertain. Examples (16) and (17) will illustrate.

In (16), the writer introduces an anecdote about a time he was walking across a busy intersection, and an old woman passing him said loudly that his zipper was open. At that moment, although he did not hear the word clearly, he immediately recognized her admonition. Inaudible portions of the woman’s utterance are marked with a question mark:

(16) *Sanjo kawaramachi koosaten no tokoro o aruiteiru to mukoo kara kobashirini yattekita obahan ga surechigaizama ni ookina koe de itta. “Ma (?) ga aite masse” (?) no tokoro ga hakkiri-shinai. Doomo maDO no *yoona* ki mo suru. Aruiwa maE datta no kamoshirenai* (Kaze no oto: Sounds of Wind)

> 3rd street name intersection LNK place ACC walking when over there from trot came old woman SBJ pass each other as loud voice P said Ma (?) SBJ open is (?) LNK part SBJ clear.NEG Somehow window LNK *yooda a feeling also have* Otherwise front was LNK not sure

‘When I was walking through the (bustling) intersection at Third Street, an old woman came up to me and said loudly, “Your (?) is open!” as she passed me. I heard something like “Ma (?) is open.” The sound (?) is not clear. I somehow felt [yooda] it could be maDO [‘window’]. Or it could have been maE [‘front’].’

Even though the inference is based on what he heard, in other words an eyewitnessed event, he uses *yooda* in company with *ki ga suru* ‘have a feeling,’ an expression that does not express definitive judgment. Accordingly, this instance indicates that such meager firsthand information is by no means established evidence all the time. Despite prevailing assumptions of the evidential theory in which firsthand information is tied to the authenticity of source information, the writer nevertheless chose *yooda over rashii*. 7

In (17) the source of uncertain information is visual, not aural.

(17) *Shueijo no mae wa kanarazu koe o kakete tooru. Shuei-san mo teineini aisatsu o kaishitekurete, kimochi ga ii. Hantaigawa no tsuuro o aruiteiru toki wa te o ageru. Garasu-goshi ni kyoshu shiteiru no ga mieru *yoona kigasuru ga, yoku wa wakaranai.* (Kaze no oto: Sounds of Wind)
Security booth **LNK** in front of **TOP** necessarily say hello walk by guard too courteously greeting **ACC** return so feeling **SBJ** good Other side **LNK** hallway **ACC** walking when **TOP** hand **ACC** raise Glass-through **LOC** salute holding **SBJ** seen **yooda** have a feeling but well **TOP know.**

‘Whenever I walk by the security guard’s office, I say hello to him. He always politely returns the greeting to me, so it makes me feel good. When I am walking on the other side of the hallway, I wave. I have a feeling [**yooda**] that I can see him saluting beyond the surveillance window, but I am not sure.’

The word **yooda** does not encode the speaker’s conviction with regard to the inferential utterance. Instead, the essayist here expresses the low quality of the source information on which his inference is based. Although he has relevant information for his inference, he does not wish to focus on it, owing to its imperfection. He straightforwardly shows his uncertainty by adding the phrase **yoku wa wakaranai** ‘I am not sure.’ This utterance is significantly contradictory to the prevailing account for **yooda** as an indicator of the speaker’s conviction.

### 4.3 Inference non-established: **daroo** vs. **soo-da**

Makino and Tsutsui (1994), Maynard (1990), and others note that **daroo** and **soo-da** are used when the speaker does not make his conjecture based on any specific information, but rather, on his own subjective thought or opinion. According to the hypothesis proposed here, **daroo and soo-da** share the meaning **inference non-established** (though they differ in the degree of focus they signal). That is, **daroo and soo-da** are used when there is no relevant source of information.

#### 4.3.1 **Daroo: Inference non-established – High Focus**

Makino and Tsutsui (1994), Maynard (1990), and Kikuchi (2000) indicate that **daroo** conveys the speaker’s mere guess, and that the speaker is not responsible for the accuracy of his conjecture. However, there are ample instances in the collected data in which speakers use **daroo** to indicate confidence in their conjectures. For instance, in (18), the writer expresses her strong conviction toward her own proposition with the verb **shinjiru** ‘believe in’.

(18) Donoyoona fukoo ga kite mo, kanojo nara shikkarito sore o kokufukushi, kesshite jiko o miushinawanai **daroo**. Okinawa no itami o jibun no itami ni kasanete sodattekita K-ko san da. … sono itami kara umareta tsuyoi chikara o shinjiru.
Any kind of misfortune SBJ came EMP she P securely it ACC overcome. and never oneself ACC get lost. NEG daroo name GEN pain ACC self GEN pain LOC overlap. and grew. came name SFX COP the pain from born strong power ACC believe

‘No matter what kind of misfortune visits her, she will certainly overcome it. She should never lose her way. K-ko-san grew up experiencing Okinawa’s agony as her own agony…. I believe [daroo] that her strength was fostered through her pain.’

Likewise, in (19), an experienced physician immediately dismisses his interlocutor’s impertinent suggestion by presenting an inference based on his expertise.

(19) “Nyuin shita hoo ga, …” to watashi ga moushiage tara, “Iya, sono hitsuyoo wa gozaimasen deshoo.”s (Shayo: Impoverished Aristocracy)

Be hospitalized (would be better) SBJ QT I SBJ SAY.HUM then No it necessity TOP Hor, NEG daroo

‘As I said, ‘I suppose that it would be better if (she were hospitalized…).” “No, I don’t think [daroo] it’s necessary.”’

Example (19) is found in a novel in which the narrator is the adult daughter of the referent, and the speaker is a physician. Despite the use of the honorific expression hitsuyoo wa gozaimasen ‘is not necessary’, the physician decisively denies the layman’s interference. He shows his professional opinion that the patient does not need to be hospitalized. Considering the social classes of the participants in this exchange, the physician’s response must be understood as a very definite reply.9

4.3.2 Soo-da: inference non-established – low focus

Unlike daroo, soo-da is used in the context of a speaker commenting on an object or event over which he exercises no direct control. In addition to many cliché expressions, soo-da also appears when a speaker is certain that his conjecture will never be realized, as in (7), above. Moreover, the speaker himself tends to be a passive participant lacking agentivity; in most cases the semantic roles are experiencer, patient, or beneficiary. Prototypical examples of soo-da include Kono sushi wa uma-soo-da ‘This sushi looks tasty,’ and Kushami ga de-soo-da ‘I feel like I’m about to sneeze.’

For instance, example (20) shows that the series of honorific verbs and affixes indicates that the patients who agree to volunteer for the medical experiment will be doing the hospital a favor. In the passage, that the hospital is asking a favor from its patients is clearly expressed with the linguistic devices of honorifics: itadake, the potential form of the humble verb ‘receive’; the honorific suffix kanja-san ‘pa-
tient-polite suffix’; the deferential verbs o-kikininaru ‘listen to us’, and o-kimeninaru ‘decide’. Since the hospital is the beneficiary in this particular case, they have to rely on their patients’ goodwill to participate in the project.

(20)  Touin de wa chiken o okonatteimasu. Chiken ni sankashite itadake-soo-na kanja-san wa … setsumei o yoku o-kikininate… o-kime kudasai (Poster at Tokyo University Medical: August 4, 2002)

This hospital LOC TOP medical experiment ACC do.be.POL experiment DAT participate HUM.receive-soo-da patient HON.suffix TOP explanation ACC well HOR.listen HOR.decide please

‘This hospital is involved in experiments to develop new medicines. If you are willing to participate in the project and cooperate as a patient with us [soo-da], please listen to your doctor’s explanation very carefully and make your commitment to the project.’

The excerpt in the actual poster is immediately followed by the following sentence:

*Chiken eno Gorikai o o-negai itashimasu.*

Medical experiment p HOR.understanding ACC HUM.plead HUM.do

We humbly ask you to understand the value of the medical experiment.

In this sentence, there are three sets of honorific expressions. First, the abstract noun rikai ‘understanding’ is prefixed with the honorific go. Second, the verb negau ‘plead’ is modified with the humble prefix and the humble form itashimasu ‘I beg you.’ Moreover, the entire word gorikai ‘your understanding’ is printed in red. The use of red ink evidently suggests that the writer (that is, the hospital administration) wants to highlight the notion ‘we are asking for your understanding of the value of the project,’ and explicitly appeals to the patients for their goodwill. The hospital may not wish to emphasize the word itadakeru ‘we are able to benefit from your assistance.’ For that reason, the LOW FOCUS soo- is located here.

Examples (21) through (23) are instances in which the soo-da inference appears with the expression konkyo wa nai ‘without basis.’ The writers plainly indicate that they are merely surmising. Their inferences are insignificant information even to the readers of their logs. Thus the writers opt for soo-da because of its meaning, that is, because the inference is not based on any relevant information and is thus presented in a reserved manner.

(21)  Kachidashi-soo-na ki ga suru n desu yo ne. Konkyo nai kedo. (Online)

Win.begin soo-da have a feeling COM COP p p basis absence but

‘I have a hunch [soo-da] that (the horse) will start a winning streak. Although I have no grounds [for this assumption].’
(22) Konkyo mo naku, nanika ga okori-soo-na ki ga suru. (Online)
Basis EMP without something SBJ happen (soo-da) have a feeling
‘Without any basis I have a premonition [soo-da] that something will happen to the world.’

(23) Shikashi, itsuka soko ni tadoritsuke-soo-na ki ga suru. Konkyo wa nai ga ...(Online)
However sometime there at reach.POT soo-da have a feeling basis TOP non but
‘I am in a disposition [soo-da] to be able to achieve that point one of these days. Although there is no basis for saying this…’

5. Conclusion

We have proposed that the structure of the Japanese inferential system is comprised of two parameters: type of source information and degree of focus. Our hypothesis is that the forms signal the speaker’s different degrees of focus on the proposition of his or her inferential judgment. The meaning of rashii and yooda are respectively high focus and low focus in the inference-established category, and the two different focuses (high vs. low) in terms of forcefulness are relative between the forms. Daroo and soo-da do not demand established information, so the speaker uses other source information such as empirical knowledge of the world. The speaker’s choice of one form over the other depends on the degree of focus he wishes to put on an inference. This study posits that the speaker’s role in the presentation of inferential judgments is much more important than acknowledged in the analyses based on epistemic modality and evidentiality in which the speaker’s word choice is tightly governed by evidence type.

Notes

1. Translations are those of the author, unless otherwise noted.
2. This discourse pattern jitsu wa + rashii is neither a coincidence nor an idiosyncrasy on the part of the speakers, but it is a widespread trend among young speakers of Japanese.
3. It is agreed that daroo (and its polite form, deshoo) is the only genuine affirmative conjecture auxiliary in modern Japanese (Saeki 1993 etc.).
4. It is said that a speaker can use daroo when making an irresponsible statement (e.g., Kikuchi 2000).
5. This is illustrated in Kirsner (1993), which includes an explicit comparison of a Columbia School analysis with a Cognitive Grammar analysis of the same grammatical category.

6. Some researchers define the morpheme soo as the suffix that creates a stative verb (i.e., keiyoo-dooshi in Japanese nomenclature). For instance, Kitahara (1981) notes that the so-called inferential auxiliary soo-da is actually an adjectival suffix. Therefore, in his analysis, Kitahara excludes the suffix soo-da from the auxiliaries. The present paper distinguishes the suffix soo-da from the hearsay soo-da as respectively, inferential auxiliary and hearsay auxiliary, following the conventional meaning.

7. The instance also invalidates Hayatsu’s (1988) claim that yooda indicates the speaker’s psychological nearness toward the source information.

8. Mio (1995 [1942]) notes that deshoo is a variant of daroo.

9. The referent in (19) is an aristocrat who lost her social status because of post-war social changes.

10. Soo-na is the attributive form of soo-da;

Abbreviations

ABL ablative case particle  
CMP complementizer  
COP copulative verb  
DAT dative case particle  
EX exclamation  
HUM humble form  
LNK linker  
NEG negation  
PAST past tense  
POT potential form  
Q interrogative marker  
SBJ subject marker  
TOP topic marker  
ACC accusative case particle  
COJ conjunctive particle  
CP communicative particle  
EMP emphatic particle  
HOR honorific form  
INF inferential auxiliary  
LOC locative case particle  
P particle  
POL polite form  
PSV passive morpheme  
QT quotative  
SFX suffix

Data Sources

Chapter 13. The structure of the Japanese inferential system


References

Chapter 14

Structuring cues of conjunctive *yet*, *but*, and *still*

A monosemic approach

Charlene D. Crupi
Rutgers University

This study examines over 400 tokens of conjunctive *yet*, *but*, and *still* to confirm predictions tied to a unique structuring cue or meaning posited for each form (Crupi 2004). The research is conducted within a Columbia School sign-based linguistic framework; however, unlike forms typically examined in CS analyses, *yet*, *but*, and *still* do not comprise a closed grammatical system. Rather the three are independent linguistic units that share the common communicative function of adversative conjunction. By examining the flow of information through a text, this research indicates that conjunctive *yet*, *but*, and *still* contribute unique and consistent clues about overall textual structure.

1. Introduction

This study represents a portion of a more extensive research project that proposes a single, constant semantic value or meaning for the English word *yet* (Crupi 2004). While *yet* serves a variety of syntactic functions in written texts, the present discussion is limited to those contexts where *yet* links one piece of information to another, traditionally termed a conjunctive use. The research is conducted within a Columbia School (CS) sign-based linguistic framework where *meaning* is defined as a value relationship between grammatical signals sharing a common semantic domain (Diver 1995; Reid 1991; Huffman 2001). However, unlike the linguistic forms typically examined in CS analyses, *yet*, *but*, and *still* do not comprise a closed grammatical system that exhaustively categorizes a clearly defined semantic domain. Rather the three are independent linguistic units that share a common communicative function – that of adversative conjunction. The goal in this analysis is to discover what factors distinguish those conjunctive contexts with
yet from those with but and still. Data supplied by the pragmatic interpretation of the larger text serve as the basis for hypothesizing distinct semantic values for yet, but, and still. This approach to word meaning is based on the assumption that communicative purpose is built right into the structure of language (Diver 1995; Schiffrin 1986). By looking beyond the confines of linked conjuncts to the larger communicative concerns of a text, this research indicates that yet, but, and still contribute unique and consistent clues about the thematic relevance of conjoined information and overall textual structure.

2. Research background

The idea that yet, but, and still might provide unique and constant structuring signals runs counter to most commonly held notions about the meanings of conjunctions. The meaning of a linking word is traditionally tied to the relationship that exists between conjoined informational units. The degree to which the conjunction itself is held responsible for all nuances within this relationship determines whether a monosemic (single meaning) or polysemic (several related meanings) approach is pursued. A sparse view limits the conjunction to its function as a logical operator (Grice 1975; Horn 1972). Willis and Dabbs (1966: 137) sum up the logical relationships that exist between elements linked by but in the following way: “But shows the general logical relationship of contrast, which includes contradiction, opposition, or concession.” Lakoff (1971) adds to the logical contrast between antonyms (e.g., John is rich, but Bill is poor) a second category of adversative contrast produced by a denial of expectations generated by informational content. An example of this second type of but-conjunction is John is tall but he’s no good at basketball. In this example, the contrast is not to be found in any direct semantic opposition, but rather between the inferred expectation that tall individuals make good basketball players and the statement contrary to that assumption found in the linked information.

Like Lakoff, Quirk et al. (1988) note that the logical function does not adequately cover a conjunction’s use in ordinary language. Their more expansive list of meanings for and is illustrated in Table 1:
Table 1. Meaning of and

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Example*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consequence</td>
<td>He heard the explosion and he (therefore) phoned the police.</td>
</tr>
<tr>
<td>Sequence</td>
<td>I washed the dishes and dried them.</td>
</tr>
<tr>
<td>Contrast</td>
<td>Robert is secretive and David is candid.</td>
</tr>
<tr>
<td>Concessive</td>
<td>She tried and (yet) she failed.</td>
</tr>
<tr>
<td>Conditional</td>
<td>Give me money and (then) I will help you.</td>
</tr>
<tr>
<td>Similarity</td>
<td>A trade agreement should be no problem and (similarly) a cultural exchange should be easily arranged.</td>
</tr>
<tr>
<td>Additive</td>
<td>He has long hair and (also) he often wears jeans.</td>
</tr>
<tr>
<td>Explanation</td>
<td>There's only one thing to do now—and that's to apologize.</td>
</tr>
</tbody>
</table>

* (Examples including parenthetical insertions, Quirk et al., 1988: 930.)

It is easy to see that by placing the full weight of the relationship between joined elements on the conjunction itself, Quirk et al. open the door to unlimited polysemy based on the variety of terms and conditions that are conjoined. For example, we might add a Reversed Sequence category with the example: Yes Mother, I dried the dishes, and I had to wash them as well, since Jason refused to help me.

Quirk et al. indirectly pass on to but the polysemy of and by defining but as “and yet,” where yet is listed as a contrastive conjunction. Thus but is assigned the additive function of and (e.g. not only..., but also expressions) and the further component of contrast or “unexpectedness in light of previous information” (p. 935).

Halliday and Hasan (1976: 237) also consider and an integral part of the meaning of but. However, they avoid Quirk et al.’s unlimited polysemy by allowing the constant adversative or “contrary to expectations” sense of but, yet, and however to operate at both the propositional and pragmatic levels of interpretation. They note that unexpectedness can be generated externally from the propositional content of the clauses or internally from the communicative situation itself (p. 250). By granting two domains of pragmatic interpretation (content level and speech-act level) to the adversative sense of yet, but, and however, Halliday and Hasan effectively limit the number of separate meanings required to account for the multiple uses of these conjunctions. Sweetser (1990) adds a third domain of pragmatic interpretation to those noted by Halliday and Hasan: an epistemic level. Conjunction within the epistemic or logical domain refers to a link between a conclusion and a previous sentence that contains the premise from which it is derived. She argues that it is possible to have only one semantic value for such linking words as but and and, but different interpretations or senses depending on the type of information that is joined (p. 10).
Advances in Functional Linguistics

Progressing through the works of Lakoff, Halliday and Hasan, and Sweetser, one can see that the gradual widening of the interpretive range of conjoined information contributes to a reduction in the polysemy ascribed to the linking word itself. However, the analytical scope of these researchers remains focused on the conjuncts themselves – that is, limited to (1) the informational content of the conjuncts; (2) assumptions, inferences or conclusions that can be drawn from the informational content of the conjuncts; or (3) possible domains of interpretation that pertain to the informational content of the conjuncts. None of these researchers considers the larger question of whether or not conjunctions provide information as to the thematic weight or relative importance of respective conjuncts to the discourse as a whole. They appear to accept the general assumption that equivalence at the level of grammatical structure indicates equal thematic importance at the discourse level. While some observe that in certain cases the order of linked conjuncts cannot be reversed, they attribute this apparent asymmetry to the influence of temporal sequencing, deductive precedence or the statement-comment directionality of the implied contrast, rather than to the relative importance of each informational unit to the text as a whole. Only Quirk et al. (1988: 920) ascribe a higher degree of prominence to the second conjunct due to its location after the conjunction. By limiting their analyses to the conjoined information, most have no data upon which to base any claims of relative importance to the larger text.

However, Sweetser’s observation that the multiple senses often ascribed to conjunctions are attributable to the type of information joined rather than to the meaning of the conjunction itself supports the claims of this study. Lang (2000) moves the investigation beyond the scope of the conjuncts proper by initiating the notion of backtracking in the text to understand an implied contrast. Schwenter (2000) adds a forward dimension to the analysis with a consideration of overall communicative purpose within the larger discourse. Lastly, Blakemore (1987), Brinton (1996), and Schiffrin (1986) shift the semantic domain from informational content (e.g. contrast or concession) to speaker attitudes about the relative importance of subsequent information. It is from this global perspective – the larger communicative goals of a text – that the current study derives constant and unique semantic values for conjunctive yet, but, and still.

3. Hypotheses

In order to facilitate the discussion, the following textual model will be adopted:

\[ \text{[Information A]} \quad \text{linking word} \quad \text{[Information B]} \]
Chapter 14. Structuring cues of conjunctive *yet, but, and still* 267

For example in the conjunction *Anna is a serious student but she has a fine sense of humor*, information A is *Anna is a serious student*; the linking word is *but*; and information B is *she has a fine sense of humor*:

<table>
<thead>
<tr>
<th>Information A</th>
<th>linking word</th>
<th>Information B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anna is a serious student</td>
<td>but</td>
<td>she has a fine sense of humor.</td>
</tr>
</tbody>
</table>

Phrased in terms of this textual model, the following hypotheses regarding lexical signals, their meanings and structuring cues are offered:

1. **Yet, significant contrast**: the contrast between information A and information B is thematically relevant.
2. **But, override**: abandon the assumptions tied to information A in favor of more thematically relevant information B.
3. **Still, continuation, no change**: information B is not new.

According to the above hypotheses, *yet* joins informational units A and B that contribute more or less equally to a contrast that is important to an author’s communicative goals. *But*, on the other hand, signals an asymmetrical relationship between linked conjuncts; the reader should abandon the assumptions tied to information A in favor of the more thematically important information B that follows. When the meaning of *still, continuation, no change*, is applied to a conjunctive context, the result is an *end detour* prompt. *Still* indicates that it is time to go back to what has already been said or to information generally shared by both writer and reader. Some examples that illustrate these hypotheses follow.

3.1  **But: The override effect**

Since *but* is the most commonly used of the three linking words under consideration, the discussion will begin with a simple construction in which *but* links two adjectives:

(1) For in 1861, when he was just twenty-four, James [Murray] met and the following year married a handsome, but delicate infant-school music teacher named Maggie Scott (Winchester, *The Professor and the Madman*, 1998, p. 34).

Taken out of context, example (1) would seem a perfect instance of a symmetrical or reversible use of *but*. With equal accuracy as to the state of real world affairs, Winchester could have written, “a *delicate*, but *handsome* infant-school music teacher.” Since there is no inherent semantic conflict between delicacy and beauty, one must infer that the author sees one attribute as positive and the other as nega-
tive. Within the confines of the sentence, however, it is impossible to determine whether or not Winchester places more weight on one or the other of these two attributes. But as one progresses through the text, no further mention of Maggie’s beauty occurs. The only other descriptives directly applied to her are “still-sickly and well beloved” and her inclusion in the “forlorn couple” (p. 35). Furthermore, the only sentence that grants Maggie the status of sole subject focuses on her ill health as she falls “gravely ill with consumption and was said by Harwick doctors to be unlikely to withstand the rigors of another long Scottish winter” (p. 35). Clearly from the standpoint of additional information offered in the text, Maggie’s weak constitution is a far more developed topic than her physical attractiveness.

When the overall theme of the book is taken into consideration, Maggie’s ill health takes on an even higher degree of relevance. Winchester’s book describes the writing of *The Oxford English Dictionary*. James Murray is in fact the professor of the title – the person who directed the progress of the *OED* from 1879 until his death in 1915. A self-educated genius, Murray had no formal education beyond the age of 14. It is his wife’s “delicate” constitution that brings the young Scotsman to London where he submits to a menial clerking position in order to care for her. After her death, Murray remains in London where he attends lectures sponsored by the local philological society, a connection that leads to his ultimate involvement in the *OED* and makes him a central character in Winchester’s book. Thus, both from the perspective of the immediate text and the overall theme of the book, it is clear that the attribute *delicate* which follows *but* contributes in a more significant way to Winchester’s narrative than the preceding attribute *handsome*.

The preceding analysis of *delicate* supports the notion that *but* introduces information of greater thematic relevance, and further serves to illustrate the many ways in which a text can be examined to reveal the author’s communicative purposes. However, it is not always necessary to carry out such an extensive analysis to discover an author’s intent – sometimes an author will clearly state his attitude toward conjoined information, as is the case in example (2):

(2) Black bears rarely attack. **But** here’s the thing. Sometimes they do. All bears are agile, cunning, and immensely strong, and they are always hungry. If they want to kill you and eat you, they can, and pretty much whenever they want. That doesn’t happen often, **but** – *and here is the absolutely salient point* – once would be enough (Bryson, *A Walk in the Woods*, 1998, p. 16, emphasis added, CC).

This passage contains two *but*s, the first introducing the fact that bears do occasionally attack and the second emphasizing the gravity of the situation when one does. Bryson’s two side comments (“here’s the thing” and “here is the absolutely salient
point” leave no doubt that author places more importance on the potential results of a bear mauling than on the somewhat reassuring fact that this rarely happens.

In both of the previous examples, information A and information B share the same relative truth status. What differs is the importance the writer places on the linked information. My conclusion is that *but* conjunctions are always asymmetrical, since information B carries a higher thematic payload.

### 3.2 Yet: A thematically important contrast

As illustrated in the following examples, contexts where *yet* is used reveal a different structuring principle than contexts where *but* is found:

(3) Scheduling seminars this semester is becoming a *nightmare*. Everyone who wants to present a topic wants everyone else to be in attendance, *yet* there is practically no one week when everyone can be in attendance (Davis, email, 2000, emphasis added, CC).

(4) There the full horror of this cruel and fearsomely bloody struggle came home to him, suddenly, without warning. Here was an *inescapable irony* of the Civil War, not known in any conflict between men before or since: the fact that this was a war fought with new and highly effective weapons, machines for the mowing down of men – and *yet* at a time when an era of poor and primitive medicine was just coming to an end. It was fought with the mortar and the musket and the minie ball, but not *yet* quite with anesthesia or with sulfonamides and penicillin. The common soldier was thus in a poorer position than at any time before: He could be monstrously ill treated by all the new weaponry, and *yet* only moderately well treated with all the old medicine (Winchester, *The Professor and the Madman*, 1998, p. 52, emphasis added, CC).

What is most strikingly consistent about these two passages is the high thematic importance of conflicting information. In example (3), the writer describes his situation as a “*nightmare*.” Example (4) focuses on an “inescapable irony” of the Civil War, the fatal combination of modern weaponry and primitive medicine that resulted in very high casualties. In both cases, it appears that information A and information B contribute in an equal way to each of these dilemmas.

Close examination of example (4) reveals that the information contained in both conjuncts is given additional attention in the subsequent text. The first instance of *yet* in this passage contrasts information A, the beginning of the modern age of weaponry, with information B, the final phase of primitive medical practices. Both these factors contribute in a significant way to the “full horror” that
eventually leads to the insanity of Winchester’s other central figure, the madman, Dr. William Minor. The contrast is picked up again in the second conjunctive use of yet that recasts the opposition in terms of the impact of these factors on the common soldier who could be “monstrously ill treated by all the new weaponry, and yet only moderately well treated with all the old medicine.”

Interestingly, there is also an instance of but in this text. In this conjunction, information A lists the modern weaponry, “mortar and the musket and the minie ball,” and information B details the modern medicine – anesthesia, sulfonamides and penicillin – that were “not yet quite” on the scene. Here it seems “not yet quite” somewhat tempers the overriding effect of but. The modern weaponry, information A, is mentioned again in the following sentence, but with less elaboration. Graphic descriptions of “machines for the mowing down of men” or specific references to weapon types do not occur. The focus shifts to the condition of the patient. In the paragraph that follows, the text goes on to describe the gangrene, amputation, filth, and so on that faced the army field doctor, and Winchester adds, “for every one who died of wounds caused by new weapons, so two died from incidental infection, illness and poor hygiene.” The “but not yet quite” combination contributes to the ultimate interpretation that modern medicine could have overridden some of the results of the new weaponry – that the old medicine could not is reaffirmed in the final yet conjunction already discussed. What seems to be the case here is that on the local level, both the modern weaponry and primitive medicine contribute equally to the immediate theme of an “inescapable irony;” but in terms of relevance to Dr. Minor’s insanity, which is the focus of the chapter, the lack of modern medicine is a more highly developed theme than the topic of modern weaponry. So, it appears that yet is less clear in indicating the relative importance of the linked information; it merely signals that each conjunct contributes to a contrast that is important to the writer’s communicative goal. Rather than dismissing previous information, yet seems to leave it available for future development.

3.3 Still: Reconnecting with earlier information

The final examples will demonstrate how still reconnects information B with themes already presented in a text.

Example (5), like (2), is taken from Bryson’s book, A Walk in the Woods. The book is a humorous account of the author’s trek along the Appalachian Trail. Naturally, large predatory animals are a serious concern to hikers. At this point in the book, Bryson is back home in New England and about to resume his hiking in an area where the last mountain lion was supposedly shot in 1903. However, he goes on to describe in detail several claims of lion sightings that have been reported since 1983. The paragraph that contains example (5) begins with the observation
that it is not “beyond the realm of possibility that mountain lions could have survived undetected in New England” (p. 203).

(5) Even so, it is unlikely that a large cat could survive in sufficient numbers to breed not just in one area but evidently all over New England and escape notice for nine decades. Still, there was that scat. Whatever it was it excreted like a mountain lion (Bryson, 1998, p. 203, emphasis added, CC).

The initial connective phrase “even so” has the effect of lowering the importance of the preceding arguments favoring the remote possibility of feral mountain lions. His conclusion is that it is highly improbable that lions have been living and breeding in New England for 90 years. Still introduces information, the presence of lion scat, that would seem to be unexpected in light of this conclusion. If Bryson’s primary purpose was to highlight the importance of this contrast, he could have used yet; but that was not his choice. Neither does there seem to be an asymmetry in the relative importance of Information A and Information B, since the ultimate solution does not involve abandoning assumptions tied to either. Bryson concludes:“The most plausible explanation was that any lions out there – if lions they were – were released pets, bought in haste and later regretted” (p. 203). What then prompts the use of still in (5)? The most plausible clue is the presence of anaphoric reference contained in information B, “that scat.” That implies that at some prior point in the text this particular scat has been mentioned. Indeed this link refers back to Bryson’s earlier description of an incident in 1994 when a Vermont farmer summoned a state wildlife biologist after seeing three lions in his backyard.

The animals were gone by the time the biologist arrived, but he found some fresh scat, which he dutifully bagged up and dispatched to a US Fish and Wildlife Laboratory. The lab report came back that it was indeed the scat of Felis concolor the Eastern mountain lion (pp. 201–202).

The factor favoring the choice of still over yet in this example is the fact that the information connected is not new. Still reconnects the reader to information already presented and then set aside in favor of a subsequent theme.

The backtracking function of still is readily observable in example (6), which appeared in a May 2000 Time magazine article, “Letter from Vietnam,” describing John McCain’s eighth visit to Hanoi, the site of his 5 ½ year incarceration as a prisoner of war:

(6) For the Hanoi government, McCain is not a source of infatuation, but he is a significant figure – and sometimes an irksome one. His outburst last week that “the wrong guys won” wasn’t exactly diplomatic. Still McCain was instrumental in pushing the U.S. to normalize relations with
its former enemy. In Hanoi, he used his clout to argue for a free-trade agreement with the U.S. that has been stalled in recent months. And he attended a “repatriation ceremony” – the transfer to American custody of the recently discovered remains of U.S. servicemen. “It is part of reconciliation,” says McCain. (Carney, p. 8)

The paragraph that precedes this passage describes the U.S. media’s “endless fascination” with the details of McCain’s personal story. The above paragraph begins with the contrasting information that the Hanoi government does not share a similar infatuation, but their disenchantment is overridden by the fact that McCain continues to be a “significant figure” in U.S.-Vietnamese relations. The author then tacks on to McCain’s significance the additional component of irksomeness. The fact that this information is set aside by a dash and linked with and contributes to the interpretation that this is a side comment, something added on to the main flow of the exposition. The subsequent sentence further develops the irksome theme by describing a specific instance when McCain irritated the Hanoi government. However, when the author wishes to signal an end to this thematic detour, he employs still to return to his main point, that of McCain’s significance in normalizing U.S. relations with Hanoi.

4. Predictions

The previous discussion highlights a variety of textual features that can be directly tied to the meanings proposed for conjunctive yet, but, and still. However, the evidence that most readily lends itself to objective analysis is the flow of information in prior and subsequent text. The pattern of informational flow most closely associated with each conjunction is as follows: Yet will be found in contexts where both information A and information B are repeated in the subsequent text. This follows from the hypothesis that yet joins conjuncts that contribute more or less equally to a contrast of thematic importance. As a result, one would expect to have both information A and B receive further comment in the subsequent text. In contrast, the proposed override function of but favors contexts where information A is not mentioned at all, or given very little elaboration in the text that follows the conjunction. As the informational component of greater thematic importance, information B, on the other hand, should be reiterated and further developed in the subsequent text. The constant semantic value continuation, no change predisposes conjunctive still for those contexts where information B has occurred previously in the text. These distributional predictions can be formalized in Figure 1:
Chapter 14. Structuring cues of conjunctive *yet*, *but*, and *still*  

- *Yet* will occur more frequently than *but* or *still* in contexts where both information A and information B are repeated in the subsequent text.

  \[
  \text{A yet B .... A .... B}
  \]

- *but* will be found more frequently than *yet* or *still* in contexts where information A receives little elaboration (-ELAB) or no additional mention; and information B is highly elaborated (+ELAB) in the subsequent text.

  \[
  \text{A but B .... B}
  \]

- *still* will be favored over *yet* and *but* in contexts where information B has been mentioned at some prior point in the text.

  \[
  \text{B .... A still B}
  \]

Figure 1. Distributional predictions

5. Data

These predictions have been confirmed through quantitative analysis of over 400 conjunctive tokens taken from two independent corpora: my own collection, the Crupi Corpus (ten complete books, plus journal, newspaper and magazine articles) totaling over 1,000,000 words of text; and the Brown Corpus, an online database of 1,000,000 words of text purportedly representative of American English. In all, three separate conjunctive samples were culled from the Crupi Corpus: a *yet* sample, a *but* sample, and a *still* sample. A fourth sample comprised of conjunctive *yets* taken from the Brown Corpus was subjected to the same quantitative methods in order to compare the results generated by the *yet* data found within the Crupi Corpus with those produced by a corpus compiled to represent American English.

6. Analytical tool

The analytical tool designed to measure the textual indicators associated with these predictions is a spreadsheet that includes a separate column for each of the textual
features to be examined. Each row represents a single occurrence of conjunctive *yet*, *but*, or *still* with scores relative to:

1. Prior occurrence of information B in text
2. Repeat of information A in subsequent text
3. Repeat of information B in subsequent text

Table 2, a simplified version of the actual analytical spreadsheet, illustrates how individual entries were logged.

Table 2. Sample spreadsheet entry, conjunctive analysis

<table>
<thead>
<tr>
<th>Link</th>
<th>Information A</th>
<th>Information B</th>
<th>Prior B</th>
<th>Repeat A</th>
<th>Repeat B</th>
</tr>
</thead>
<tbody>
<tr>
<td>yet</td>
<td>The death of a man is unique</td>
<td>it is universal</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Since all conjunctions are coded relative to all features (that is, *still* is coded for repeat A and B, as well as its own test feature, prior B; *yet* and *but* are coded for prior B, as well as those associated with their own proposed meanings, repeat A and B), the spreadsheets effectively test and provide control data for all three linking words.

7. Results

Table 3 gives the results of the conjunctive analysis of *yet*, *but*, and *still*. As indicated in the first column, the number of tokens within each sample varied across the three conjunctions under investigation (e.g., N = 107 for Yet/BC, that is, the *yet*s taken from the Brown Corpus; N = 128 for Yet/Crupi). For the sake of comparison, then, the numeric values in Table 3 (and all subsequent tables) represent percentages of the total number of conjunctive tokens for that sample. Additionally, each general “Repeat” column is followed by “+Elab” and “-Elab” columns, which divide that “Repeat” category according to the degree of elaboration of the repeated information. “Repeat” tokens that merited more than two words of text were ranked “+Elab” as compared to those “Repeats” that received minimal attention in the subsequent text (two words or less; an indirect reference) which were rated “-Elab.” Under each of the “+Elab” and “-Elab” percentages for the total sample is given the percentage for the repeat subcategory.
Table 3. Percentages of prior and repeat information in conjunctive contextual model

<table>
<thead>
<tr>
<th>Linking Word</th>
<th>Prior B</th>
<th>Repeat A (+) Elab</th>
<th>Repeat A (–) Elab</th>
<th>Repeat B (+) Elab</th>
<th>Repeat B (–) Elab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yet/BC N = 107</td>
<td>19</td>
<td>80</td>
<td>71</td>
<td>9</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>88</td>
<td>12</td>
<td></td>
<td>97</td>
</tr>
<tr>
<td>Yet/Crupi N = 128</td>
<td>16</td>
<td>85</td>
<td>70</td>
<td>15</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>83</td>
<td>17</td>
<td></td>
<td>92</td>
</tr>
<tr>
<td>Still/Crupi N = 82</td>
<td>94</td>
<td>65</td>
<td>54</td>
<td>11</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>83</td>
<td>17</td>
<td></td>
<td>97</td>
</tr>
<tr>
<td>But/Crupi N = 128</td>
<td>23</td>
<td>24</td>
<td>10</td>
<td>14</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42</td>
<td>58</td>
<td></td>
<td>95</td>
</tr>
</tbody>
</table>

The first two rows of Table 3 reveal a striking similarity between the yet data of the Brown Corpus and the Crupi Corpus. In fact, the correlational coefficient between the two yet samples is .995, an indicator of a very strong positive linear association between the two samples. The high correlation between these two samples confirms the reliability of the yet data provided by the Crupi Corpus. As a result, the findings for yet are generalizable to the degree that one accepts the Brown Corpus to be representative of standard English usage. While the analytical support for the observed contextual effects of but and still is limited to the Crupi Corpus, the close correlation between the two samplings of yet would imply a high degree of reliability for these results as well.

Table 3 confirms all three predicted contextual patterns. To facilitate the presentation of data, these predictions will be discussed one by one, accompanied by breakout tables with the pertinent data highlighted.

Prediction 1: Yet will occur more frequently than but or still in contexts where both information A and information B receive additional elaboration in the subsequent text.

The hypothesized meaning of yet, significant contrast, implies that both conjuncts contribute equally to the featured contrast. As a consequence, one would expect that both components would receive further mention in the text; and that, in general, the repeat rates for information A and information B would be roughly equivalent for yet. Table 4 highlights the repeat rates for information A and B, as well as the difference between the two (repeat B – repeat A) for all three lexical items.
A quick comparison of the Repeat A column and the Repeat B column of Table 4 reveals a notable difference in the range of variation between the two. Across the four samples analyzed, Repeat B percentages vary by only 10 points (88% for both yet samples and still, 98% for but). By contrast, Repeat A percentages listed in the first column vary from 85% (yet Brown Corpus) to 24% for but, a range of 61 percentage points. The difference between Repeat B and Repeat A rates for each of the linking forms serves as one indicator of the difference in relative thematic importance between the conjoined informational units. This difference is calculated in the final column of Table 4. For yet/Brown Corpus the difference between the repeat percentages is 8% (88% – 80%); for yet/Crupi Corpus the difference is only 3% (88% – 85%); as compared to a 23% difference for still (88% – 65%) and a 74% difference for but (98% – 24%). The two yet samples clearly demonstrate the lowest differences (8% and 3%) between Repeat B and A percentages; that is, in comparison to the tested environments where but and still are found, yet occurs in passages that exhibit the least asymmetry in thematic relevance between linked conjuncts. These findings confirm the prediction that yet is used to link conjuncts of relatively equal importance to the overall communicative goals of a writer. At 80% and 85% repeat rates for A and 88% repeat rates for B, the data indicate that yet is usually situated in contexts where both A and B receive additional attention in the text, thereby confirming the first prediction.

Prediction 2: But will be found more frequently than yet or still in contexts where information A receives little (-Elab) or no additional elaboration; and information B is highly elaborated (+Elab) in subsequent text.
Table 5 lists the prior and repeat percentages for *but*:

**Table 5. Percentages of prior and repeat**

<table>
<thead>
<tr>
<th>Linking Word</th>
<th>Prior B</th>
<th>Repeat A (+) Elab</th>
<th>Repeat B (+) Elab</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>but</em></td>
<td>23</td>
<td>24</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>58</td>
<td>5</td>
</tr>
</tbody>
</table>

Consistent with the *override* hypothesis, Table 5 reveals that *but* is used to link conjuncts of unequal importance to a writer’s communicative goals. The lower thematic importance of information A is clearly reflected in its percentage of repeats, only 24%, the lowest for Repeat A. Of the Repeat A sub-category for *but*, 58% are characterized as -Elab (low elaboration); that is, the repeat is just a single word or two or an oblique reference. The 58% -Elab versus 42% +Elab represents the only skewing in favor of low elaboration among the ratings for all other lexical items in both corpora. The total percentage of +Elab (highly elaborated) information A repeats for *but* stands at just 10% of the total sample, a strong indication of its reduced relevance in the subsequent text.

The predicted high degree of thematic importance of information B is robustly confirmed by its 98% repeat rate, the highest among all features tested across all three lexical items. The +Elab information B repeat rate is 93% for the entire *but* sample, and 95% of the Repeat B sub-category. The 74% difference (see Table 4) between the repeat percentages of information A and B offers convincing support for the structuring value posited for *but*: abandon assumptions of information A for the more thematically relevant assumptions of information B.

**Prediction 3:** *Still* will be favored in contexts where information B has been mentioned at some prior point in the text.

Table 6 highlights the information relevant to the discussion of *still*-conjunctions.

Table 6 indicates that prediction 3 accurately describes those contexts that feature *still*. 94% of the total instances of conjunctive *still* are situated in contexts where information B has been given explicit prior mention in the text, in contrast to 23% for *but*, and 19% and 16% for the *yet* samples. What is noteworthy about the percentages for *still* is that the rate for Prior B (94%) surpasses both the repeat rates for A (65%) and B (88%).

Table 6 reveals a gradual shift in the relative importance of information A and B across the lexical items under investigation. The data present a scalar effect across the three lexical items. Initially, the shift from *yet* to *still* produces a reduc-
tion in the importance of A, while the relevance of B remains constant. As one progresses from still to but, the repeat percentages show a simultaneous increase in the importance of information B (up to a 98% repeat rate) and a precipitous drop in the relative importance of information A (down to a 24% repeat rate). Still effectively occupies a middle position between the relative equivalence that exists between conjuncts linked by yet and the heavily weighted difference in thematic relevance exhibited by informational units conjoined by but.

Table 6. Percentages of prior and repeat information in conjunctive contextual model

<table>
<thead>
<tr>
<th>Linking Word</th>
<th>Prior B</th>
<th>Repeat A (+) Elab</th>
<th>Repeat A (–) Elab</th>
<th>Repeat B (+) Elab</th>
<th>Repeat B (–) Elab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yet/BC</td>
<td>19</td>
<td>80 (–) 12 71</td>
<td>9 (–) 12 88</td>
<td>88 (–) 9 85</td>
<td>3 (–) 3 97</td>
</tr>
<tr>
<td>Yet</td>
<td>16</td>
<td>85 (–) 17 70</td>
<td>15 (–) 17 83</td>
<td>88 (–) 7 81</td>
<td>7 (–) 8 92</td>
</tr>
<tr>
<td>Still</td>
<td>94</td>
<td>65 (–) 17 54</td>
<td>11 (–) 17 83</td>
<td>88 (–) 3 85</td>
<td>3 (–) 8 97</td>
</tr>
<tr>
<td>But</td>
<td>23</td>
<td>24 (–) 58 10</td>
<td>14 (–) 58 42</td>
<td>98 (–) 5 93</td>
<td>5 (–) 5 95</td>
</tr>
</tbody>
</table>

To summarize, Tables 3 through 6 offer quantitative support for the predicted differences in contexts where yet, but, and still perform linking functions. The validation of these predictions, in turn, provides evidence that yet, but, and still perform distinct roles in structuring texts – roles that can be accounted for by the hypothesized meanings and structuring cue: yet, significant contrast; but, override; and still, continuation, no change.

8. Conclusion

The findings of this study indicate that while yet, but, and still are all used at points of disjuncture within texts, each provides distinct instructions to the reader as to how subsequent information fits into overall textual structure. Yet indicates that the contrast between prior and subsequent information is important to the larger themes of the text. Still returns the reader’s attention to information already present in the text or to shared knowledge of the larger context. But signals that the information that follows is of higher importance than preceding information. The override structuring effect of but, supported by the quantitative analysis of data taken
Chapter 14. Structuring cues of conjunctive *yet, but, and still*

From over two million words of written text, challenges the commonly held notion that *and* is part of its core meaning (Halliday and Hasan 1976; Quirk et al. 1988).³

Within the tradition of Columbia School analysis, this study represents an important step in the application of the sign-based research model to the lexicon. As Diver noted, the difficulty of applying to the lexicon the Saussurean model that worked so well for the grammar is that relationships and oppositions between words in the lexicon overlap and crisscross in ways that defy neat division (Diver 1995). Consequently, most CS research has dealt with tightly constructed grammatical systems whose limited members exhaustively characterize a given semantic domain (e.g., Diver 1969; Garcia 1975; Huffman 1997, 2002; Reid 1991; but see also Davis 1995; Goldberg 1995; Stern, this volume). Those CS studies that do address lexical questions have, in effect, treated lexical oppositions as quasi-grammatical systems (*at, on, and in* in Reid 2004; *only and just* in Tobin 1995). While de Jonge (1993) does not formally propose that the semantic domain ‘to resemble’ is exhaustively covered by the Italian verbs *parere* (to indirectly resemble) and *sembrare* (to directly resemble), the opposition he investigates is limited to two members. The present study has adapted the CS analytical model to accommodate the loosely connected semantic network of the lexicon. The fact that *but* and *still* exhibit oppositions with other lexical items not included in this study is a strong indicator that *yet, but,* and *still* are not tightly defined by their relationship with one another as are the components of grammatical systems. What this study has shown is that each word signals an independent meaning that overlaps in predictable ways with the meanings of the others. By extending the CS research model beyond closed grammatical systems, this study demonstrates that forms as ostensibly polysemous as *yet, but,* and *still* can be reduced to a single semantic value.

**Notes**

1. Their example of an external adversative is: “The total came out wrong. *Yet* all the figures were correct; they’d been checked.” In this example, the element of unexpectedness derives from the informational or logical contrast between wrong and correct (p. 252). An example of an internal adversative is taken from Lewis Carroll’s *Through the Looking Glass*: Alice to the cat: “... you might catch a bat, and that’s very like a mouse, you know. *But* do cats eat bats, I wonder?” Halliday and Hasan note that the denied expectation in this example is Alice’s intention to give useful information; *but* precedes her afterthought that perhaps what she is saying isn’t helpful at all. Here, the contrast is generated within the context of the speech situation (Halliday and Hasan 1976: 253).

2. Since this research did not include all uses of *but, override,* as indicated by italic type, has not been formally elevated here to the level of a meaning for *but, override* is the proposed structuring cue of conjunctive *but.*
3. According to Blakemore (1987: 139), *and* indicates that the combination of the two conjuncts is of higher thematic relevance than the two conjuncts considered independently. Since *but* clearly indicates that the second conjunct is of higher thematic importance than the first, *but* is semantically at odds with the idea that the connection is more important than either conjunct considered independently.

Data Sources

Davis, Joseph C. M. 2000. Email correspondence sent November 11, 2000 to participants of the CSLS seminar.

References

Chapter 14. Structuring cues of conjunctive *yet, but,* and *still*


CHAPTER 15

The case for articulatory gestures – not sounds – as the physical embodiment of speech signs

Thomas Eccardt
Independent Scholar

The term *articulatory gestures* is common among linguists, amounting to a kind of analogy with the manual gestures of sign language. This paper takes the term seriously, rejecting the notion that sounds are the physical embodiment of the linguistic sign. Making the case for the gesture as a legitimate type of sign, it shows how vocal movements are far more convincing candidates for the signifiers of human language when viewed from several different semiotic perspectives, including physiology, physics, psychology, and communication theory.

1. Introduction

It is puzzling why so many linguists use the term *articulatory gestures* to refer to the vocal movements of speech, when the general assumption is that the *sounds* are the signs of language. Perhaps it is even more puzzling that sounds are considered signs at all, given that sound and light are only the means by which humans perceive gestures, such as articulatory gestures or those of American Sign Language. The purpose of this paper is to make the case for designating the articulatory gestures to be the signifiers of spoken language and to explain how the paradoxical acoustical view arose. First, it will make the case for *gesture* as a legitimate category of *sign*. Second, it will differentiate between the gestural *symbol* and the acoustic *signal*, both traditionally considered types of *sign*. It will show that a *symbol* is a discrete, arbitrary, perceptible entity, whereas a *signal* is a disturbance of a medium through which the symbol may be perceived. In this case, the *signal* is not at all arbitrary, and therefore not a *sign*. The method will be to investigate sounds and gestures from a number of viewpoints – as listed by Hockett (1960) – some of which have already been seen to underline the importance of articulatory gestures
in speech. But this paper will take a more general, semiotic perspective, and show that these viewpoints designate gestures, not sounds, as the physical embodiment of the speech sign.

2.1 The nature of gesture

The word gesture in modern English always denotes a sign, that is, a specific kind of physical signifier (signifiant), with an associated signified (signifié). A gesture is never simply an action, as gestum meant in Latin, but any meaningful action. Nor is a gesture merely the result of an action. For example, a hotel guest might make the comment “what a nice gesture,” on finding a chocolate on her freshly made bed. She's not referring to the chocolate itself; she means the placing of a chocolate on her bed. The chocolate (or the visual image of the chocolate) is merely a physical trace of the placement, just as the sounds of language are the result of articulatory gestures.

Gesture has not played a major role in semiotics; typically, research has been limited to the manual and facial gestures that accompany speech or to body language. Meanwhile, sign languages for the deaf, such as American Sign Language (ASL), have been neglected by semioticians. This article may be the first semiotics study to deal systematically with the articulatory gestures of speech as real gestures. Perhaps the reluctance to deal with gesture is due to its temporary, fleeting nature. A gesture lasts as long as the time it takes to produce it, no longer. Results or traces of gestures, such as a chocolate on a bed, may be more permanent. Other traces of gestures might include the arrangement of artworks in public areas and the arrangement of chessmen on a board as the result of various moves. In each case, the meaningful element, the sign, is the action, not the result of the action. So yet another categorization of signs seems useful: temporary versus permanent signs. Temporary signs (gestures) would correspond to Hockett’s (1960) “rapid fading” in his list of dimensions useful in comparing human speech to the communicative behavior of other species. Table 1, found before Section 3.3, gives an extended list, based on Nöth (1990: 156).

But not all actions in semiosis are gestures. Especially in the production of permanent signs and non-replicable signs, the actions necessary to create such signs are rarely gestures. For example, the hammer blows on a chisel used in the production of a statue could hardly be called gestures. Although the non-gestural movements involved in the production of signs may not always resemble brushstrokes, it will be useful to call them strokes to distinguish them from gestures. Now the central question in this paper can be reformulated in somewhat more precise terminology:
Are the articulatory movements of speech merely the strokes required in the production and transmission of sound symbols? Or are the articulatory gestures the signifiers of speech, perceived mostly through the medium of sound?

Now, if articulatory gestures are to be considered gestural signs in a true semiotic sense, then each signifier must correspond to only one signified. So an articulatory gesture must consist of an entire morpheme, despite the popular tendency to call a single movement an articulatory gesture. Of course, each speech sign is usually made up of more than one such movement or subgesture, just as written words are made up of multiple letters. Such components of signs are known more generally as figurae in semiotics. The problem of what constitutes an individual subgesture – also known as “segmentation” – will be disregarded here unless it becomes important (cf. Davis, this volume). Whether a subgesture corresponds to a phoneme, something less than a phoneme, or something more than a phoneme is just as irrelevant as the question of whether a signal flag gesture consists of the action of one arm, two arms, or something more than that. The point is that the sign consists of the flag gestures, not the flags, and especially not the light reflected by the flags.

2.2 Other definitions of gesture

Armstrong et al. (1995: 42) have remarked “a growing number of researchers are attempting to explain speech (but, oddly, not yet [ASL] signs) as gesture.” This paper defines all gestures as signs, since they are signifying actions or deeds. It follows that articulatory gestures could constitute the physical aspect of the signs of speech. But this is not the first time that articulatory gestures have been recognized as signifiers. The psychologist George Mead (1934: 46) called them “vocal gestures,” classifying them as “significant symbols,” and considering the gesture to be an individual act of social behavior. Mead’s mentor, Wundt (1897: 299), had a similar conception of the gesture, and saw the vocal gesture as having evolved from the “natural” gestures of sign language. Perhaps because gesture has never been recognized as a separate species of sign, these earlier assertions appear to have been forgotten.

Unfortunately, the characteristic iconicity of many types of gesture has led to much confusion on the nature of gesture and its proper definition. Wundt (1897: 299) considered sign language inferior, erroneously claiming that “signs for abstract concepts are entirely wanting.” Jakobson (1971: 318) calls certain non-linguistic onomatopoetic sounds gestes vocaux. He gives the example of /ööö/ for the sound of a fire truck made by a child whose language does not include the /ö/ vowel. Grammont (1946: 413) uses the term gestes articulatoires to denote iconic morphemes whose mouth shapes resemble the things they represent. Some linguists even define gesture as any non-linguistic communicative action. Almost
Advances in Functional Linguistics

entirely in this spirit, the most common classification of gestures has arisen: systems of visible gestures (where they are substitutes for words, such as “thumbs up” for “okay”); kinesics (or gesticulation, the gestures that accompany speech); pantomime (and the gestures of dancing and acting); body language (and other gestural indications of emotion); and finally sign languages (such as American Sign Language or Plains Indian Sign Language). Excluded in the common definition of gesture are articulatory gestures and all other meaningful actions not involving visible body motions, such as the sending of flowers.

2.3 The gesture as a sign

In the language of flowers, when one receives pennyroyal, the message is to “flee away.” Here the sign is not flowers, but the gesture of sending flowers. Observing pennyroyal growing in the wild gives no cause to flee. Nor should the florist flee when she receives a shipment of pennyroyal from her hothouse. A wise spy would not flee, either, if somehow he knew that the florist mistakenly sent him pennyroyal, when his cohort actually had ordered periwinkle. Only the gesture of sending (or ordering) pennyroyal is a meaningful sign. This author has a practice of putting an empty container of any product on the kitchen table to remind himself to buy more of the item. Here again the sign is not the item, but the gesture of placing it on the table. The empty container has no significance in its ordinary place in the household; on the table it becomes the trace of a meaningful gestural sign. These gestures leave permanent traces, and do not have to be perceived immediately. Others, like articulatory gestures (whose acoustical traces normally fade rapidly), must be perceived immediately. A sound recording of speech, on the other hand, might be counted as a permanent trace of articulatory gestures.

The gesture of placing an object in an unusual spot also recalls Baron’s (1981: 23) fundamental method of referring, namely, “presentation” (cf. Eco’s (1979: 224) “ostention”). By picking up and holding out a toy ball, one may be referring to all balls, for example. Baron claims that the ball would be an iconic sign, “re-presenting” all balls. But surely this is as much a gesture as pointing to a ball. It may be better to call this sign “the gesture of presenting a ball,” rather than “a ball presented as a sign” because once the sample ball is put down, it is no longer a sign. The gesture of presenting or moving an object also creates a kind of disturbance of the environment, in the sense that the object is no longer in its normal position, calling attention to the fact that it has been moved. The sound waves which make the articulatory gestures perceptible are also considered disturbances in the medium of the air.
2.3.1 Specialization

Another characteristic of linguistic signs – and perhaps all signs – is Hockett’s “specialization.” A semiotically specialized sign must not chiefly serve some other utilitarian function, otherwise it is merely a symptom. Hockett’s example is the female stickleback fish, whose abdomen is distended when filled with unfertilized eggs. Although it also signals the male that she is ready for mating, the swelling is chiefly due to the eggs, and is used only secondarily for communication. Such symptoms, when recognized by inference, are not even signs, according to Eco (1979: 17), unless they become conventionalized in a society. Nöth (1990: 395) also points out that “while there is little doubt about conventional gestures being signs, the semioticity of practical behavior has been questioned.” So actions, too, can be specialized and semiotic, or simply utilitarian. When an object is created, it may be mostly utilitarian, as in the creation of a clay pot, or it may be mostly semiotic, as in the sculpting of a clay statue. When an object is moved, it may be moved for utilitarian purposes, as in the transport of goods, or it may be moved for semiotic reasons, as in the manipulation of signal flags, the sending of flowers, and the vibration of air in speaking. When an object is destroyed, this may be done for a practical purpose, as in demolition, or it may be destroyed as a symbolic act, as in terrorism. Like other semiotic notions, including the rapid fading of gesture, all these semiotic characteristics appear in various gradations, and are not necessarily mutually exclusive categories of signs, as Sebeok (1994: 21) notes, in agreement with Eco.

3. Semiotics and theoretical methods

Hockett’s (1960) sixteen dimensions are intended to compare human speech to other forms of animal communication, but many of them parallel the traditional characteristics which semioticians have found in the sign, as shown in the “semiotics designation” column of Table 1 below. This paper is organized according to Hockett’s dimensions, which seem to be the widest-ranging classification of communicative behavior. Each dimension offers a different perspective from which to examine the question of sound versus gesture. Although Hockett does not address the relevance of semiotics to linguistics, this paper will. Following Sausser, linguistics will here be considered a branch of semiotics, and not the other way around. This is because semiotics is the study of signs, which are the tools of communicative behavior in general, and language is a form of human communicative behavior in particular. Accordingly, evidence from semiotics will be used to show that articulatory gestures, rather than sounds, are the true signifiers of speech. Hervey (1982: 7) considers the best method for developing semiotics to
be that of comparing various semiotic systems, and this comparative method will
be employed here. Evidence will be sought from fields outside semiotics, such as
psychology and communication theory. In fact, there are various schools of pho-
etics whose major point of agreement lies essentially in considering articulatory
gestures to be the phonological units of speech.

3.1 Linguists’ acoustical views of speech signs

It is surprising, then, that those few linguists who considered themselves semioti-
cians almost unanimously found the linguistic signifier in the sounds of speech,
even as they differed on the nature of the meaningful part of the sign. The father of
modern linguistics and semiology, Saussure, considered the linguistic sign to be a
psychological entity that united a sound image with a concept. Saussure (1966: 11)
even claimed that in speech production, a concept first “unlocks” a sound image,
then the brain transmits a corresponding image to activate the speech organs. This
seems to imply that speakers mentally hear their own speech before they produce
it. Saussure was reacting against traditional linguists who “neglect the auditory
side” (1966: 38). In a footnote (1966: 66), Saussure’s editors explain his acoustical
bias through his belief that language was something “received from without.” But
elsewhere Saussure (1966: 8) says: “we simply cannot reduce language to sound or
detach sound from oral articulation.”

Some semiologists expressed no opinion other than to quote Saussure direct-
ly.¹ Perhaps Saussure’s influence was so strong that many semioticians simply took
sound as signifier for granted. In his expansive Theory of Semiotics, it appears that
the only time Eco (1979: 247) even indirectly mentions the physical embodiment
of linguistic signs comes in the middle of his chapter of sign production when he
discusses “certain sound parameters which permit the recognition of the replica.”
Diver (1979) was one of the few to express a contrary opinion (Note: Diver’s term
signal corresponds essentially to signifier, and gesture to subgesture; more on termi-
nology below):

The overall interrelationship among phonology, physiological phonetics, and
acoustic phonetics can be sketched as follows: the speaker learns the signals of the
language (the morphemes) as made up of a limited number of distinct articulatory
gestures; these are the phonological units. In the particular circumstance of the
individual acts of speech, the attempt at the articulatory gesture produces certain
vocal movements. These can be recorded and observed on, for instance, X-ray film.
The vocal movements in turn shape and excite resonant cavities, and the resulting
sounds can be recorded and analyzed with the spectrograph. The movements and
sounds, then, are consequences of the articulatory gestures. (Diver 1979: 181)
Diver was led to his belief in the articulatory subgestures as phonological units because his investigations into the skewing of their distribution indicated only articulatory causes, not acoustic ones. On the other hand, Diver (1995: 48) takes the sounds of speech as the basic observables of phonology. Curiously, since sound is only one of several means to observe articulatory gestures.

3.2 Symbols vs. stimuli vs. signals

Although not a semiotician himself, Bloomfield did consider speech to be communicative behavior. But his conception of this behavior was heavily influenced by a stimulus-and-response viewpoint: “the gap between the bodies of the speaker and the hearer – the discontinuity of the nervous systems – is bridged by sound waves” (1933: 26). But if two brains were directly connected by sound waves, there would be no symbolic behavior, because both sounds and the electric impulses sent between nerve cells are not symbols, but stimuli. Hockett’s seventh dimension, “semanticity,” distinguishes symbols from stimuli. A symbol has only an indirect link with behavior, whereas a nerve cell will automatically fire when sufficiently stimulated and will pass on the signal to other cells with which it has connections. There may be a statistical cause-and-effect relation between the perception of a morpheme and the behaviors it evokes in the long run, but a given morpheme in general can produce various and individually unpredictable responses, even in similar contexts. Ironically, Bloomfield’s implied analogy between nerve firings and sound is quite appropriate. Sounds can be considered stimuli, which automatically evoke vibration of the hairs in the cochlea, which automatically cause the auditory nerve cells to fire. One of the few areas of agreement by almost all semioticians is the difference between a symbol and a stimulus. Sebeok (1994) distinguishes them as two of his six species of signs. But he uses an unfortunate choice of terminology, calling stimuli “signals.” Eco (1979: 19) actually considers this distinction to be the very boundary of semiotics: “stimuli cannot be regarded as signs.” Essentially, Eco thinks that every sign must be at least somewhat arbitrary, and something that physically triggers a response is certainly not that.

Many define a signal as the physical embodiment of any sign. Cherry (1966: 308) calls it the “physical embodiment of the message.” Diver’s (1979: 182) definition in the above citation seems to fit this description. But semiotics already has a term for the physical embodiment of a sign, namely, signifier. In communication theory, the signal is the physical phenomenon (e.g. a sound wave) which travels from a source to a receiver, carrying information. Noise (static) often interferes with this signal as it travels through a medium, such as the air.
Table 1. Hockett's dimensions for comparing communicative behavior in different species.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Channel</td>
<td>(Light, sound, etc)</td>
<td>X</td>
<td>.</td>
<td>.</td>
<td>5</td>
<td>Channel</td>
<td>Multisensory</td>
<td>Unisensory</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Transmission and Reception</td>
<td>(How message travels)</td>
<td>X</td>
<td>.</td>
<td>.</td>
<td>6</td>
<td>Transmission</td>
<td>Multidimensional</td>
<td>Unidimensional</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Rapid fading</td>
<td>(Gesture, temporary)</td>
<td>X</td>
<td>.</td>
<td>.</td>
<td>2</td>
<td>Gesture</td>
<td>Static or dynamic</td>
<td>Dynamic</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Interchangeability</td>
<td>Com-signs, replicability</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>9</td>
<td>Interchangeability</td>
<td>Gestures or strokes</td>
<td>Strokes</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Total feedback</td>
<td>(Can be monitored by sender)</td>
<td>X</td>
<td>.</td>
<td>X</td>
<td>5.2</td>
<td>Feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Specialization</td>
<td>Not a symptom (not utilitarian)</td>
<td>X</td>
<td>.</td>
<td>X</td>
<td>7</td>
<td>Production</td>
<td>Uttered / specialized</td>
<td>Emitted / symptomatic</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Semantics</td>
<td>Not a stimulus (indirect link w/ behavior)</td>
<td>X</td>
<td>.</td>
<td>X</td>
<td>8</td>
<td>Perception</td>
<td>Interpreted</td>
<td>Reacted to</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Arbitrariness</td>
<td>Not an icon</td>
<td>X</td>
<td>X</td>
<td>.</td>
<td>11</td>
<td>Iconicity</td>
<td>Non-motivated or motivated</td>
<td>Motivated</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Discreteness</td>
<td>(Not continuous)</td>
<td>X</td>
<td>.</td>
<td>.</td>
<td>4</td>
<td>Discreteness</td>
<td>Discrete</td>
<td>Continuous or discrete</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Displacement</td>
<td>Not an index</td>
<td>X</td>
<td>X</td>
<td>.</td>
<td></td>
<td>Independence from signified</td>
<td>Independence from signified</td>
<td>Nearness to signified</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Productivity</td>
<td>Sememes (syntagms)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Duality</td>
<td>Figurae (double articulation)</td>
<td>X</td>
<td>.</td>
<td>.</td>
<td>6.3</td>
<td>Dimensionality</td>
<td>Double or single articulation</td>
<td>Single articulation</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Cultural or traditional transmission</td>
<td>(Subject to non-genetic change)</td>
<td>.</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>12</td>
<td>Language evolution</td>
<td>Cultural / societal or instinctual</td>
<td>Instinctual</td>
</tr>
<tr>
<td>14</td>
<td>Prevarication</td>
<td>(Can be false/meaningless)</td>
<td>.</td>
<td>X</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Reflexiveness</td>
<td>(Can talk about itself)</td>
<td>.</td>
<td>X</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Learnability</td>
<td>(Foreign language can be learned)</td>
<td>.</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>10</td>
<td>Acquisition</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Information loss</td>
<td>(Not a dimension)</td>
<td>X</td>
<td>.</td>
<td>.</td>
<td>6.4</td>
<td>Distortion</td>
<td>Predictable distortion</td>
<td>Random &quot;white&quot; noise</td>
<td></td>
</tr>
</tbody>
</table>
This definition of *signal* corresponds in a way to the *traces* in the gestural examples that have been mentioned in this paper, and it is the one adopted here. Signals seem to be simultaneously both symptoms and stimuli, because they are the direct result of the actions of the sender and they also stimulate the sensors of the receiver. For example, the sounds (stimuli) that hit your ear are the result (symptoms) of someone else’s articulatory gestures. The flowers that you discover at your doorstep are the result of the actions of the delivery person. This kind of signal or trace is like a function whose output unfailingly corresponds to some specific input. So it also seems that there is a positive correlation between Hockett’s specialization and his semanticity. That is, symbols are typically specialized and therefore semantic, while signals or traces are neither specialized nor semantic.

Sometimes one signal is transformed into another. At a radio station, the sounds of speech are transformed into radio waves, which are then transformed back to sound waves by the receivers in the listeners’ homes. These radio waves are a direct analog of the sound waves, and in a formal sense they have just as big a claim to be called the signifiers of speech as do sound waves. Therefore, if the signifier is by definition to be found in only one part of the speech act, it cannot be in a signal.

### 3.3 Hockett’s dimensions

Table 1 lists, one per row, Hockett’s sixteen dimensions for comparing communicative behavior in different species. The last row (without a number) is not really a dimension, but is added to show how information loss enters the picture. The first two columns list the dimension number and Hockett’s name for each. The third column explains the dimension and/or gives its traditional semiotic category. The next four columns show how each dimension is defined in terms of semiotic participants. For instance, in the tenth row, under “displacement” or the semiotic term “not an index,” both the signifier and signified columns contain an “X.” This shows that an *index* is defined in terms of its signifier and signified, namely that there must be little or no displacement between the signifier and the signified. For example, a road sign is an index because it must be placed near the road it indicates. The next two columns give the section number and section name within this paper where the dimension is discussed. Because this paper confines itself to the physical embodiment of signs, there are sections to cover almost all the dimensions with an “X” in the signifier column, and sections to cover only a few without an “X.” Observe that among the rows with an “X” in the signifier column, several do not have traditional semiotics designations (in bold type) in the third column. This may indicate that semioticians have not paid enough attention to the physical embodiment of signs.
Two traditional semiotic species of signs, *signals* and *symbols*, do not appear as rows of this table. This is because they represent instead a fundamental dichotomy in the physical embodiment of signs. Symbols are usually defined in terms of what they are *not* – a symbol has no motivated link with what it denotes (or anything else, for that matter). So if there were a row in the table for symbols, there would be no “X” in any of columns 4 through 7. Being completely independent, conventional, and arbitrary, the symbol is the sign *par excellence*. On the other hand, *signals*, as defined above, have a very strong connection between the sender and the receiver. In fact, the signal *is* the link between the sender and receiver. As such, a pure signal is not really a sign at all. The last two columns of the table illustrate how Hockett’s dimensions are typically realized by symbols versus signals (or traces). Notice that many of the dimensions can be realized in only one way by either aspect of the dichotomy – for example, symbols are always discrete, even though signals can be either continuous (e.g., a sound wave) or discrete (e.g., digital mobile phone transmissions). This fundamental dichotomy between symbols and signals distinguishes two opposing kinds of communication: communication through *behavior* versus *mechanical* (e.g., electronic) communication; for example, communication between sentient creatures through linguistic symbols versus communication within creatures through nerve signals.

4. **Discreteness vs. continuity**

A continuous signal cannot be unmotivated, because it varies continuously in intensity as a function of something else. If it is divided into energy levels, these levels can be assigned arbitrary meanings, but then the signal is effectively discrete. Conversely, a discrete signifier can be unmotivated, i.e., arbitrary. And linguistic signs are typically unmotivated. So if it could be shown that articulatory gestures were more discrete than speech sounds, the case for gestures as signifiers would be strengthened. Saussure, on the contrary, apparently believed that segmentation (discrete categorization) could be done algorithmically on sounds, and not with a moving picture of the gestures. But he had never seen either an x-ray film of the gestures or a sound spectrogram. It is a question for further research as to which is more discrete and which is more continuous: mouth movements or spectrographic changes.
Unlike most of Hockett’s other dimensions, the choice of channel is neither binary nor scalar. Whereas a sign is either mostly iconic or mostly symbolic, humans have five different senses or channels through which to perceive that sign. There is also a kinesthetic (muscular) sense, which comes into play in speech feedback. Any one of these could be used as a communications channel. But as noted in the last two columns of Table 1, symbols can often be detected through multiple channels, while signals are usually transmitted in only one. This is because a symbol is a kind of physically existing discrete thing, even when it’s a gesture. An example would be a statue, which can be seen, felt, smelled and even tasted. When a metal statue is struck, it might produce a ringing sound. On the other hand, a signal, such as light or sound, can be detected only through one organ, either the eye or the ear. Although sign languages are normally perceived through the visual medium, Helen Keller used her sense of touch to perceive the manual gestures of fingerspelling. Spoken language gestures are perceived through four senses: hearing, sight, touch, and the kinesthetic sense (the latter two in feedback only). Sound, being only one of the signals through which gestures are perceived, cannot be considered the signifier.

5.1 Lip reading

Perhaps due to the influence of audio electronic media such as radio and telephony, most linguists have simply ignored the visual channel of speech. Diver (1995: 110) was once again the exception, in that he recognized the relevance of “vision to the extent that lip-reading plays a role in the phonology.” Nor do ordinary speakers of language ignore it, and they often find themselves staring at their interlocutor’s mouth. An American politician once said, “Read my lips, no new taxes!” His audience understood the metaphor: This is so important that if you can’t understand it through the acoustic channel, use the visual channel. It is well known that people with mild hearing loss compensate through lip-reading. (Cf. Davis 1984 [1987]: 43–45, which is more explicit than Diver on the role of visibility.)

Much experimental evidence reveals the importance of the visual channel in speech. Summerfield (1987: 5–6) summarizes the evidence showing that speech understanding in the presence of noise improves greatly when the listener can observe the face of the talker. The “McGurk effect” shows how much people are willing and able to rely on visual evidence in speech perception: when a recorded voice says /ba/ but a simultaneous face image says /ga/, ninety percent of subjects regularly report hearing /da/. It doesn’t even matter if the subjects are told how they are to be tricked, the visual input still affects their perceptions in the same way. If the sound were the signifier, people ought to be able to ignore visual cues
and pay attention only to the acoustical sign. Kuhl and Meltzoff (1997: 22) report that the McGurk effect works even when the voice is female and the face is male. Dodd (1987: 165–16) notes that infants as young as three months are able to detect an experimentally produced incongruity between lip movements and speech. And Mills (1987: 155) reports that “sighted children learn those sounds that have a visible articulation more quickly than those that do not have such a clearly visible articulation,” and furthermore that “visually handicapped children, who cannot observe lip movements, are slower in their acquisition of these sounds and make a different pattern of errors than do sighted children.”

5.2 Feedback mechanisms

One’s own speech is perceived chiefly through the senses of hearing and touch, plus the kinesthetic sense. Occasionally one’s own lips are visible by looking directly down towards the mouth. The acoustic and especially the visual feedback is different from what is perceived by others, while the tactile and kinesthetic senses are not available when perceiving other people’s speech. ASL signers, too, get tactile and kinesthetic feedback, and their visual feedback is even more distorted than acoustic speech feedback. Articulatory gestures are usually perceived simultaneously by speaker and interlocutor only through sound – perhaps this is another reason for linguists’ bias towards sound. But the fact that only light is common to signers and their interlocutors doesn’t seem to affect our willingness to consider ASL a system of gestures, rather than light symbols (cf. Sections 6.2–6.3).

5.3 Multisensory perception

Given that so many senses take part in the perception of speech, and given the importance of lip-reading, it would seem unwise to assign the linguistic sign to sound. In the “speech chain” (Denes and Pinson 1993: 5) there is a chain reaction starting from motor neuron firings, continuing through articulatory muscle movements, articulatory gestures, sound waves, vibrations in the ear, and ending more or less at firings of the auditory nerves. Other reactions, including mental (brain) reactions, are less predictable. There is another speech chain for the visual perception of speech, and two more chains (tactile and kinesthetic) for feedback. Starting again with motor neuron firings, each chain has every link in common with the other three up to and including the point of the articulatory gesture; thereafter the chains diverge. Traced backwards, these speech chains converge at and point to the articulatory gestures as the true signifiers of speech.

Some have even claimed that the categorization of perceived objects actually depends on multisensory perception. Armstrong et al. (1995: 50) summarize
Edelman’s Theory of Neuronal Group Selection as requiring “at least two separate sensory channels which supply signals to neural maps.” It seems that most discrete entities we identify, whether symbolic or not, are perceptible through more than one channel. Given how messy spectrograms are, it is plausible that infants categorize and “segment” the acoustic input they hear with the help of as many senses as are available, including feedback.

6. Transmission and reception

The audio and visual channels are special because of their high channel capacity (amount of information per time), and their speed of transmission. Chemical messages in the body, for example, are much slower than sound and electromagnetic radiation. Nor can chemicals be used to return information about the positions of objects in the environment or their shapes. Sound and light can. But there is a major difference between echolocation and most vision. In echolocation, a dolphin, for example, produces its own high-frequency sound to reflect off the objects around it, whereas most creatures take advantage of the sun’s light or other illumination for vision. And this is the major difference between the transmission of speech and sign language. In speech, the shapes of the other articulators are detected through the sounds produced by the vocal cords. In sign language, all gestures are detected through external illumination, and signers can’t communicate in the dark.

6.1 Sound production vs. sound modification

Morris (1946: 33) accepts the conventional view that sign language gestures are signs, but articulatory gestures are not. He recognizes that some signs (such as visually perceived gestures) are activities, but he considers the sounds of language to be products of activities. Accordingly, whenever such products exist, the products, not the activities, are the signs. This is somewhat odd, since light from Morris’s visual gestures (which can be considered photons) is more of a concrete “product” than sound (which is only vibration). He was also probably unaware that the vast majority of the distinguishable differences in speech sounds are due to modifications made by the motions of articulators other than the vocal cords. In other words, most phonetic distinctions result from sound modification, not sound production. All vowels, semivowels, semiconsonants and nasals are distinguished exclusively by the resonance caused by the shape of the mouth and/or the nasal cavity. Stop consonants make a small almost inaudible noise, and they, too are distinguished by resonance occurring in the transition to neighboring vowels.
Only fricatives and affricates make their own relatively loud noise, but they, too, are distinguished mostly through neighboring vowels. In English, only /s/ and /ʃ/ are easily distinguishable by their own sound, created by different slit shapes for escaping air, whereas /f/ and /θ/ require accompanying vowels to be distinguishable from each other. On the other hand, the vocal cords do their own signaling whenever they pause briefly for a glottal stop or an unvoiced consonant, and also during the production of tones in tone languages. However, this information imparted by the vocal cords must be less than essential, because speech through a continuous monotone mechanical larynx is generally comprehensible.

Even absent an actual phoneme frequency count for Spanish, it is still possible to estimate the relative contribution of sound modification versus that of sound creation for this language by using the letter frequencies in texts and the nature of sounds they represent, since Spanish spelling is highly phonemic. Data from a survey of Spanish letter frequencies (Fontanillo: 1986) were recalculated, based on the above description, counting unvoicing subgestures plus 50% of fricative subgestures as sound production -- all other subgestures were counted as sound modification. The recalibration reveals that by letter frequency Spanish is 18.1% sound creation and 81.9% sound modification. By information contribution of the letters, Spanish is 15.5% sound creation and 84.5% sound modification. Tone languages would probably have a higher percentage of sound creation, but certainly not nearly fifty percent. On the other hand, the vocal cords produce sound through the action of closing tight during exhalation, so Bhaskararao (1994: 74) can say that “every speech sound produced in continuous speech results from the modification of the vocal tract shape.” From this point of view, speech can be considered 100% gesture. Speech is somewhat unusual, in that it is comprised of gestures plus the means (i.e., sound signal) of perceiving those gestures. In this sense, it is difficult to find another natural communication system to compare to it. Bees dancing outside the hive rely on the sun to illuminate their gestures and produce a light signal, whereas animals that produce pheromone signals produce no gestures. If one were to insist that any energy transmission at all constitutes sign production, then obviously speech sounds, not gestures, would have to be the signs. But supposing humans were equipped to produce their own (blinking) light when they use sign language, then one would also have to claim that sign language is not gesture, but luminous sign production.

6.2 The spatial configuration of the speech chain

A dissenter could make the argument that the internal production of speech sounds is irrelevant: what comes out of the mouth are sound symbols. This argument has two problems. First, articulatory gestures are not entirely internal – the lips are ex-
ternal, and sometimes the tongue protrudes out of the mouth. Secondly, the internal / external question appeals to the visual bias of “seeing is believing.” If no sounds were produced in speech, and if the human vocal tract were transparent so that speech could be completely “lip-read,” we would be much more willing to consider speech to be a form of gesturing. But because of our visual bias, what is internal and invisible also seems unobservable. However, articulatory gestures are quite observable: we hear them. Everybody knows how to make the /k/ sound, but lay people may not know that it’s made with the back of that partly visible articulator, the tongue.

Although the internal / external question is irrelevant, the observable / unobservable question is not. If signs are the tools of communicative behavior, then only behaviors that an organism can ordinarily observe are relevant. The speech chain begins with muscular neuron activity and continues with articulator muscle contraction, but these processes are not directly observable and cannot be considered signs. As it happens, the first ordinarily observable link in the speech chain, articulatory gestures, constitute the physical embodiment of the sign. Yet partly because of its central position in the speech chain, linguists have assigned sound to the physical embodiment of speech signs. The central position of sound in speech is just as irrelevant as the central position of light in sign language.

6.3 Dimensionality / double articulation

ASL signs are articulated in a three-dimensional space, whereas most articulatory subgestures can be described by only two dimensions: their degree of constriction and their position up and down the vocal tract. Each of these types of gesture loses a dimension as it is reflected into its light or sound signal. Our retinas provide us with a two-dimensional image, whereas our ears give us only one (frequency). This dimensional advantage of light allows much greater information transfer and is undoubtedly related to our “seeing is believing” visual bias. Then how is it that speech and ASL seem to transfer information at about the same rate? The double articulation of speech is the answer, as it allows a limited number of subgestures (phonemes) to signal a huge number of gestures (morphemes) in a very short time. So the double articulation might be called the third dimension of articulatory gestures.

The bias for the sense of sight probably barred the use of much iconicity through articulator shape. The articulatory mouth shapes, largely hidden from the eye but perceived by sound, touch, and the kinesthetic sense, cannot easily be connected with the shapes we normally perceive by eye in the outside world. Furthermore, we identify most things by sight, not by sound, so despite our ability to produce a great variety of sounds, we do not make much use of sound iconicity, either.
6.4 Distortion

As stated above, a three dimensional sign language gesture gets compressed onto a two-dimensional retina. This camera-like solution for perceiving the direction of light also has a drawback: distortion. The two-dimensional image of a statue or of a sign language gesture can vary considerably depending on the location of the viewer. Although the form remains the same, some aspects change their apparent shape as the viewer moves around the object viewed, and some features even disappear as they are hidden behind others. Although distortion and noise both result in the loss of information, distortion – but not noise – is regular and can be predicted. Another difference is that under distortion, in many cases there is no preferred viewpoint, and no way to avoid information loss, because all aspects of the perceived objects cannot be observed simultaneously. But more noise can always be removed, and there is always one theoretically ideal reception.

6.4.1 Spectrogram variation

Because sound does not have to travel in straight lines, little distortion occurs due to the location of the listener. However there are several phenomena in speech which can be compared to the distortion of viewpoint. One is the changing nature of the auditory speech stream, easily apparent on the spectrogram. During speech, our articulators are constantly in motion, and having a consistent acoustical view of their positions could certainly be useful in speech perception – lip-readers benefit greatly from a consistent visual view of the lips and the jaw. But when the vocal cords pause for an unvoiced sound, the formants disappear, leaving either a blank spectrogram or else something of a very different nature. Fricatives produce aperiodic noise and their mouth shapes cause absorption lines as opposed to the vocalic resonant formant lines (Dew and Jensen 1977: 200). When nasal consonants are produced, we get yet another viewpoint, because now the sound comes out exclusively through the nasal cavity, resulting in a very different resonance chamber and a very different-looking spectrogram. The result is that the most striking aspect of the spectrogram besides the formants themselves is their constant interruption. Here, consistency is sacrificed for somewhat more diverse acoustic images, especially in the case of fricatives. This inconsistency may be one reason it so difficult to train even hearing people to read spectrograms on the fly: it is very difficult to trace the movements of the articulators on a spectrogram. Yet humans are well equipped by nature to “see” through these distortions when they hear speech, just as they instinctively ignore the changing visual shapes of objects when their viewpoint changes. In the case of sign language gestures or articulatory gestures, the signal may be distorted by the viewpoint or by the actions of the lar-
ynx or the velum, yet the one constant throughout is the uninterrupted fluctuating shape of the gestures.

6.4.2 Ambiguity
An effect of distortion is ambiguity. For example, in a familiar optical illusion, the stick-figure of cube drawn on flat paper allows the viewer to imagine it sticking out of the paper or alternatively carved into it. The ambiguity arises only because the drawing is a projection of a three-dimensional object onto a two-dimensional paper. Apparently sound images can have multiple articulatory gesture interpretations as well. Fowler (1994: 4202) states, “theorists have claimed that multiple vocal-tract shapes can give rise to the same acoustic signal, so that acoustic signals are not, in fact, specific to their articulatory source.” If this is true, it is because sound, like light, is subject to regular information loss – distortion – and is an imperfect reporter of the articulators’ activities. Fowler explains that although one sound can be made in multiple ways, the neighboring subgestures disambiguate them, since only one subgesture anatomically fits into the ensemble. This implies that both sound and this “anatomical sense” are needed to correctly distinguish or learn to produce articulatory gestures.

6.4.3 Vocal cord variation
Another source of distortion is vocal cord variation. There is much leeway in the permissible actions of the vocal cords in the production of any speech sign (morpheme), somewhat less in tone languages. A sentence can be whispered, pronounced in a creaky voice, sung, or intonated in a great variety of ways, always remaining the “same” sentence. Even inhaled speech is comprehensible. These varieties can look remarkably different on a spectrogram, yet the only difference is in the performance of the vocal cords – all other articulators must make the same gestures.

6.4.4 Gestural overlap
It is well known that there is much overlap among the articulatory subgestures in speech – “segmentation” is truly a misnomer (cf. Davis, this volume). This “co-articulation” probably adds redundancy to the resultant sound, although it seems to be mostly the result of the anatomy of speech production. It is easy to see this in ASL, which also has coarticulation. And in spoken language, articulatory subgesture overlap increases when talking speed increases, just as it does in ASL. Although the increase in overlap is not uniform between all kinds of articulatory subgestures, it is largely predictable, and, according to Browman and Goldstein (1992: 163) it can explain much “allophonic variation.” Browman and Goldstein (1992: 172) also note that overlap can become so great, even in normal speech, that the main part of certain subgestures can be overlapped by others. In an expression
such as perfect memory the sound of the final /t/ in perfect can be swallowed up by the production of the previous /k/ and the following /m/. The speaker actually makes the entire /t/ subgesture, even though only a small co-articulation, the approach of the tongue to the alveolar ridge, is heard. This occultation of sounds is reminiscent of the obscuration of some visual features of a statue by others due to the location of the viewer. While the shape of the statue and the gesture are untouched, the signal reaching the eye or ear is distorted and information is lost. So gestural overlap can be considered another form of articulatory distortion.

6.4.5 Inter-speaker variation
Although a true distortion usually implies the loss of information, a linear transformation does not. So if the only difference among the speech of men, women, and children lay in the pitch of their speech due to the size and stiffness of the vocal cords and/or the size of the vocal tract, one could not claim that the acoustic transformation of one voice to another involves distortion. However, because of the complicated shape of the vocal tract, “the relationship between speech spectra and the vocal tract shapes that produce them is highly non-linear” (O’Shaughnessy 2000: 88). And as Kuhl and Meltzoff (1997: 12) point out, computers cannot as yet be programmed to recognize the similarities of the sounds produced by the same gesture among various speakers. So even if the articulation or modification of the sound is done in the same way by all speakers from infancy to adulthood, the resultant sound output is not related by a simple linear transform. Therefore inter-speaker variation can be considered another form of distortion.

6.4.6 Summary of distortion
So there are at least four types of distortion characteristic of speech: spectrogram variation, vocal cord variation, gestural overlap, and inter-speaker variation. Despite the resultant sound variations in the reproduction of a given morpheme under different circumstances due to distortion, the gesture is consistently produced more or less in the same way and can be described in the same way (cf. Section 11). Distortion amounts to a kind of break in the tight speech chain, allowing a single subgesture to produce several different sounds and allowing a single sound to be produced by multiple gestures. Other links in the chain are more solid: the same sequence of muscle movements will produce the same gestures, and the same sounds will produce the same vibrations of the hairs of the cochlea. The result is that articulatory gestures cannot be considered stimuli, although sounds can. Gestures must be counted as symbols, and as such, they must be actively perceived by listeners, rather than reacted to by hearers. This is one reason why automatic speech recognition is so difficult, while telephonic sound transmission is not.
7. **Speech production**

Is speech psychologically gesture production or sound production? This is an ongoing debate among phoneticians, with experimental evidence seeming to point both ways. Much of it hinges on whether the “target” for any speech unit is articulatory or acoustic. Then when subjects are somehow hindered from achieving their targets, do they compensate acoustically or gesturally? For the acoustic point of view, Gabioud (1994: 224) cites evidence that when jaw openings are restricted, people compensate with different lip openings to produce the same vowel sound. For the opposite (articulatory) point of view, Fowler (1994: 4201) cites analogous evidence that when the jaw is prohibited from closing, people adjust their lips to complete the consonantal closure subgesture. But compensation under duress seems like a weak argument: would it make any difference whether ASL signers compensated by making their experimentally inhibited gestures look like the uninhibited ones? Even if they did, they are nevertheless gesturing, not producing light signs. And just as it makes no difference what muscles signers use to produce ASL gestures, it makes no difference whether all speakers use the same motor commands to produce articulatory gestures.

More telling may be evidence relating to speakers’ normal behavior and their lack of compensation in the face of distortion. An adult male’s spectrogram is not a linear transform of the one he produced in childhood, yet his gestures are still produced in the same way. As a boy grows up, he makes no effort to retain his childhood sounds, even if he is able. When men need to verify which morpheme a woman has uttered, as they repeat it they do not adjust their speech to imitate the way the woman said it, by raising their pitch, for example. On the other hand, a parrot’s “speech” is in imitation of one particular speaker, or perhaps a few distinct speakers. A parrot is not gesturing in the same sense as people do – it is not imitating gestures, it is imitating sounds.

8. **Speech perception**

Deacon notes:

> we tend to perceive speech sounds in terms of ‘articulatory gestures,’ whose boundaries and distinctions correspond to articulatory (i.e. somatomotor features), not just sound features, and this perceptual process is mostly automatic, taking place beyond the access of conscious analysis. (Deacon 1997: 359)

Several schools of phonology subscribe to this theory, that there is an unconscious processor that reconstructs the gestures from distorted sounds. The motor theory
of speech perception states that in parsing a sound signal, a listener makes use of her knowledge of the relationship between her own gestures and her own output. Although this theory has fallen somewhat out of favor, others such as the “articulatory phonology” and the “direct realist theory,” have emerged, with the common thread of considering the speech units to be gestural, not acoustical. A gestural image in the brain seems more plausible than Saussure’s idea of an acoustic image. If the model for perception is a sound, then whose sound is it? Is it one’s own voice? That would imply that men should understand men faster than they understand women, and conversely. For a given morpheme or even a given vowel, what intonation lies behind the acoustic model?

According to Diver (1995: 111), “the hearer is constantly confronted with the task of identifying the intended phonological unit on the basis of all available information.” This is certainly a better orientation than a strictly acoustic or articulatory view of perception, given the evidence of lip-reading. Although speech can be understood exclusively through the sound channel, claiming that speech signs are sounds is a little like saying that footsteps are sounds. Walking can be detected by ear or by eye or even by both simultaneously. Like speech, one’s own footsteps can be detected also by the tactile and kinesthetic senses. Somehow the brain coordinates the observations of all the senses into a coherent perception of footsteps, rather than perception of sounds or light. The footsteps are perceived as a remote object, not as stimulus at the ears or eyes.6

9. Interchangeability and replicability

Hockett lists interchangeability as one of his dimensions, which is equivalent to Morris’s category of “com-signs.” Eco is concerned with replicability of signs, but if both sender and receiver can interchangeably create and receive the sign, then the sign is automatically replicable. It does seem that if a sign is interchangeable, we are more willing to call it a symbol, not just a symptom or stimulus. For example, since every vervet monkey both receives and sends alarm calls (which vary according to the danger involved), it seems the monkeys must be sending a conscious message. If a cat hisses, it may be sending a kind of message to its enemies, but the meaning must vary according to the species of the recipient. Without a consistent signified, this action can hardly be called a symbol. Furthermore, if only a sculptor is able to produce the strokes necessary to create a statue, then the ordinary viewer cannot consider those strokes to be gestures, since the viewer would probably not even know what strokes were needed to produce it.

Of course, humans produce and detect both sounds and articulatory gestures. However, because of inter-speaker sound variation, the sound is not the same. In-
stead, the articulatory gesture is the target. When we veer too far from the target, our acoustic, kinesthetic, and/or tactile feedback tells us we have made a speech error, and we correct it. That way, all speakers normally produce a given gesture similarly enough for all of us to consider it the same thing. What we all have in common is the ability to produce non-distorted gestures, plus the ability to “un-distort” the sound they produce and to interpret them in the same way.

10. Language Acquisition

First and second language learners aim to learn what gestures are needed to produce the unfamiliar sounds of the new language. Foreign language textbooks that give impressionistic descriptions of sounds rather than gestures are a thing of the past. Sometimes people listen to language tapes, hoping to match their sounds with those of the speaker. But how can a woman learn “a new sound” from the recording of a male speaker? She does so with her built-in ability to recognize matching gestures no matter who produces them. The same is true of child language acquisition. A baby needs to be able to hone its own speech to that of its parents: when they match the child can stop honing. But children's speech sounds can never completely match their parents. If they didn't have the ability to match many varieties of speech sounds, then firstborn children would have a handicap compared to secondborn siblings, because only the latter could take advantage of a child model closer to their own. Kuhl and Meltzoff (1997: 12) give a useful summary of what they call “talker normalization”: the ability of even two-month-old infants to perceive similarity among sounds that belong to a particular category, even though they are discriminably different.” This paradox is resolved by realizing that infants are not perceiving different sounds, they are perceiving similar gestures.

11. Iconicity

Iconicity in language has probably had a distracting effect on linguists and semioticians, causing them to concentrate on the sounds as symbols, rather than on the gestures. But even though articulatory gestures constitute the physical embodiment of the signs, a sound icon does not have to be a gestural (shape) icon as well. That is, an iconic sign does not have to be iconic in all media (channels) through which it can be perceived. In sign language, there are many iconic gestures which would have little iconic meaning to a blind person detecting them by the sense of touch. For example, for “house,” the exterior shape of a house is traced by two hands, starting with the sloping roof. A innately blind person, never having felt the
sloping roof of a real house, would not recognize the iconicity of this gesture. So this gesture is iconic only through the light image it suggests, not its shape, which is what distinguishes it from other gestures.

On the other hand, linguists have always relied on the gesture (articulation) to describe speech sounds, since ancient times. It has always been very difficult to describe speech sounds acoustically, mostly because there are very few natural sounds to compare them to. King Sejong of Korea found this to be the case in the fifteenth century when he designed the Hangul script to resemble mouth shapes, rather than sounds. So did the designers of the International Phonetic Alphabet, whose every symbol is defined by its articulation, not by its sound. Only recently has the spectrogram inspired terms such as *compact* and *diffuse* in the relatively new field of acoustic phonology. But it is still much easier to describe a /b/ as a flat lip closure than to describe its effects on a spectrogram, which vary depending on the vowels surrounding it.

12. Language evolution

Most linguistic evolutionary changes involve small changes to articulatory sub-gestures. And that usually implies small sound changes as well. But as mentioned above, the relationship between vocal tract shapes and corresponding spectrograms is nonlinear. This is particularly true at the narrowest constrictions of the vocal tract where turbulence may suddenly come into play. That is, a small gestural change from a stop to a homorganic fricative (e.g. /t/ > /θ/) has a big effect on the spectrogram. Yet fricativization and defricativization are among the most common sound changes. Just as the spectrogram of the speech stream boldly changes from resonance to aperiodic noise or to silence or to nasal resonance, so do speech communities tolerate big changes on the spectrogram in the evolution of language. What they rarely tolerate is big changes to articulatory gestures, such as the Japanese /h/ reflex of the historical Chinese /p/. This is because articulatory gestures are the physical embodiment of speech signs. When a group of people modify their gestural production of a morpheme, there is only so much leeway which the rest of the speech community will tolerate as a “close enough” gestural reproduction of that morpheme and adopt the change. Sound similarity is far less important. Of course, sound *differences* serve another purpose entirely, to distinguish subgestures and ultimately signs from one another.
13. Summary

The purpose of this paper has been to gather evidence from diverse areas and to show that articulatory gestures, not sounds, are the signifiers of human speech. It has made the case for the gesture as a category of sign. It has shown that the dichotomy between signal and symbol is important in semiotics, as it is correlated with 11 of Hockett’s 16 dimensions for comparing communicative behavior. And it has shown that as pure signals, sounds cannot be considered signifiers, but that gestures, as symbols, can.

However, it is not just an interesting observation that articulatory gestures are the signifiers of speech; it is a theory. This theory explains the power of lip reading, the McGurk effect, and why radio waves and spectrograms are not signifiers. It also makes the prediction that articulatory gestures are more likely to be amenable to computer analysis than their resultant sound spectrograms, since they are not distorted and should be more discrete. And perhaps the theory of articulatory gestures will inspire others to do cross-disciplinary linguistic studies under the discipline of semiotics.

Notes

1. Hjelmslev (1961: 50) takes a similar position to Saussure’s, calling the substance of his content plane “thought,” and that of his expression plane a “sound-chain.” Jakobson (1971: 656) is a little less mentalist, calling the signans an “opposite sound property,” while the signatum is not a thought, but the ability to differentiate meanings. Shaumyan (1987: 32) identifies phones as sounds, and language signs as being made up of phonemes.

2. In several cases, the semiotic category is given in terms of what it is not, e.g. not a symptom (row 6). This is because Hockett presented his dimensions in terms of human language, which makes little use of non-arbitrary signs, such as symptoms.

3. Kinesthetic sensations are induced by the contraction of the articulator muscles, and tactile sensations result from contact between articularators (Dew and Jensen 1977: 241).

4. Incidentally, intonation in English, being analogue rather than discrete, ought to be classified semiotically as a signal, not a set of symbols.

5. According to Schein and Stewart (1995: 40), the sign for ‘hate’ is made at eye level when it follows the word for ‘dream,’ another eye-level sign. But ‘hate’ is signed at chest level when it follows the chest-level sign for ‘eggs.’ This makes the phrases ‘to hate to dream’ and ‘to hate eggs’ faster and easier to sign.

6. As Fowler (1994: 4200) puts it: “listeners hear a sound as emanating from the location in space of the mechanical sound-producing event in the environment; they do not hear it at the ears where the effective acoustic signal is located.” Gibson (1979: 54) puts it as a kind of contradiction: “nothing can be seen but light, yet light can never be seen.”
7. Oshima-Takane et al (1996: 631) have found only one difference in first language acquisition between firstborns and secondborns: secondborns learn to use personal pronouns faster, because they are exposed to conversations between their caregivers and their older siblings. However there was no significant difference in number of intelligible utterances between the firstborns and secondborns at a given age.

8. If iconicity in speech is chiefly acoustic, at least three authors have recognized articulatory iconicity. As mentioned above, Grammont (1946: 413) classifies as gestes articulatoires such terms as French puer and puant, because their labial initials visually depict disgust. Anderson (1998: 180) cites words such as tongue and teeth as examples of “kinesthesia.” And Fischer (1999: 126) explains as articulatory iconicity the well-known tendency among languages for small or near things to be denoted by words in /i/ and big or distant things by words in /a/. “How else can they be explained?” he asks, “…since neither size nor distance involve sound.” Here, the tongue makes a close approach to the roof of the mouth in a gesture remarkably similar to the hand gesture for “tiny” performed with the thumb and the index finger. Another familiar sound icon that might be better explained by articulation is the famous /sl/ series, as in slide, slither, slip, sleigh, etc. When /sl/ is pronounced, the tongue slides forward along the top of the mouth. More tongue-related articulatory icons might be tip, touch, and tap. Spew, spit, and gag might also be considered iconic, in the way they articulatorily mime the actions they represent.

9. Basically, there are two types of changes: syntagmatic and paradigmatic. Syntagmatic changes include deletions, insertions, metathesis, lengthening, shortening, etc. These are changes that relate to the order or timing of subgestures in speech, and they typically relate to neighboring subgestures. The exceptions are phenomena such as vowel harmony, which relate subgestures separated by other subgestures. But in many cases of vowel harmony, there is so much overlap that there may not really be any distance between the two correlated vowels anyway. And that is the general principle of syntagmatic changes: there is little or no “action at a distance.” Some changes which have been considered paradigmatic are actually syntagmatic, such as voicing and devoicing, nasalization and denasalization. Nasalization is usually only the spreading of a nasal subgesture (opening of the velum) so that it becomes simultaneous with additional subgestures. Voicing is really the deletion of the subgesture that pauses (opens) the vocal cords. As for true paradigmatic changes, the principle is the same: little or no action at a distance. Changes such as vowel raising, vowel lowering, fricativization and defricativization involve small changes of the aperture of the constriction. Palatalization and fusion usually involve sounds with neighboring places of articulation, as well. Perhaps only in assimilation does place of articulation change radically, but presumably the resultant sound changes a lot as well.

References

Chapter 15. The case for articulatory gestures


CHAPTER 16

Meaning in nonlinguistic systems

Observations, remarks, and hypotheses on food, architecture, and honor in Kenya

Robert A. Leonard
Hofstra University

This article extends the boundaries of Columbia School linguistic semantic theory by applying its analytical constructs to nonlinguistic behaviors, where, as in language, there exist systematicity and arbitrariness: food; construction of social and gender identity; and use of architectural, private, and urban space. Further, meaningful elements of these behaviors vary analogously to Labovian sociolinguistic feature variation. The guiding orientation is that human behavior is structured not by an unmotivated, autonomous culture, but is communicative and social, interpreted by people as signals with meanings; and these meanings are discerned as interpreting power, prestige and identity. Data come primarily from fieldwork in Lamu (Kenya) and Thailand.

1. Background: Data collection and description

In 1991, my research associate Wendy Saliba and I had just returned from Africa and were assembling an inventory of Swahili street foods.¹ This was for a study of the social meaning of East African food behaviors we were to present at the Food Symposium at Oxford University. As we charted out the data, we saw that typical behaviors of who could consume what, and where, were strikingly non-random and highly constrained. It reminded us of the types of grammatical systems hypothesized in Columbia School language analyses.² We were struck by the apparent coordination of the interlocking systems of food, space, and societal identity, and by the way the interplay of elements seemed to cooperate in the categorization of a semantic substance; e.g., the way the English meanings one and other than one are defined in terms of each other and cooperate in exhaustively categorizing all possible meanings in the semantic substance of Number (Diver 1975; cf. Reid
1991: 82–83). Further, these elements varied in a way similar to Labovian sociolinguistic feature variation.

At Oxford we presented the interlocking occurrence rules, but without semantic analysis. Reviewing our paper (subsequently published as Leonard and Saliba 1992) in *Natural History*, Sokolov (1991: 78) referred to the “stratified and rule-bound society of Kenya’s Indian Ocean coast,” and characterized the whole set of interrelated behaviors and values as “an intricate code, deciphered by Leonard and Saliba.” But that presentation was limited to description. This paper carries that effort forward into explanation by introducing systematic meaning into the analysis.

2. **Food as an instance of l’arbitraire du signe**

Not only was the set of behaviors an “intricate code,” it appeared arbitrary. As I worked on our earlier analyses, I gradually came to realize – to my surprise – that food was arbitrary. This important point runs counter to common thought on food, and thus requires some discussion before we turn to the use of food in Lamu, Kenya.

2.1 The identity of food is arbitrary

The members of each culture eat different items. Typically, each group believes that the reason they eat what they eat is because that is food. Most other things in the world, i.e., what they do not eat, are *not* food. People, further, believe that food is the most non-arbitrary of substances: that its *identity* and its *meaning* are *intrinsic*. Yet a moment’s observation of the culture next door shows that other people eat things that to us are not foods. They digest them and thrive on them, yet we are sure nevertheless they are not food. No culture I am aware of allows, let alone suggests, its members eat every edible item available. Americans typically do not eat insects, while nationals of some other countries are disgusted by our peanut butter. So food itself is arbitrary: we do not eat, nor consider “food” all those things which are biochemically eligible for food status, but only some of those substances.

2.2 The perception and categorization of food is arbitrary

How we *perceive* those substances is arbitrary too. For example, that two “eggs” belong to the same category of things is an arbitrary, *linguistic* fact. Sitting under the hen – or indeed the hen and the snake – are two different, physically separate things. It is our language that categorizes them both as “eggs.” (See Ellis 1993: 27–44 for an especially cogent explanation of language as categorization.)
2.3 The meaning of food is arbitrary

The *valuative meaning* of foods is arbitrary, as well. For example, in New York, canned (tinned) food is marked as less good than fresh, as not enhancing the social status of the person who serves such a food to friends. The opposite is true in some Papua New Guinea societies, where tinned food announces the urbanity of the possessor and consumer.

Regarding a different food, the philosopher Susan Neiman (1992) relates a story heard in Berlin:

> In Silesia in the thirties there was a mixed Polish German population. Those who were in favor of German annexation stopped eating garlic. From one day to the next. *Heim ins Reich*. It was a movement. (Neiman 1992).

The meaning of garlic changed overnight.

Thus the identity, the categorization, and the meaning of food are all arbitrary. We might say that food – to borrow Wolkomir’s (1998) well-known phrase about American Sign Language – is “not mouth stuff, but brain stuff.”

The arbitrariness of food is the first step toward being available for analysis by Columbia School linguistic techniques. To the extent that Columbia School theory has been successful in describing how real speakers communicate in languages, it might well also help explain other meaningful behaviors that seem to be organized into systems. Perhaps too this attempt to extend the range of explainable phenomena might contribute to proofing the linguistic theory itself and might someday provide insights and modifications that would strengthen its ability to deal with meaning in language.⁶

2.4 Foods as evidence of otherness

Differences in food identity, characterization, and valuation distance groups from each other, and similarities promote internal cohesion. Like language, food serves well for contrastive self-identification (Leonard and Saliba 1999a).

The Nazis vilified garlic as a stinking non-food that the vermin, subhuman “garlic Jews” ate. As mentioned above, a Nazi group whose cuisine included garlic suddenly stopped its use. Clearly the physical reality of the garlic did not change; its meaning did.⁷ Later, in the 1970s, German youth in the student movement ate food mixed with raw garlic to show their opposition to the then German political hierarchy (Gregory D. Kershner, personal communication, 2005).

In a bar in Mombasa, Kenya, a Kikuyu man once whispered to me that people from the Jaluo tribe, the Kikuyus’ traditional enemies, were not only uncircumcised, but also actually ate fish.
The reason that the meaning of foods is so transformable, from group to group and from time to time is that the meaning of foods is arbitrary.

2.5 The changing value of food

Consider the durian (see Leonard and Saliba 1999a and 1999b). The Southeast Asian durian fruit is supremely popular, revered, expensive, and of bedrock importance in the culture. Piper (1989: 17), an expert on Southeast Asian fruits says, "It is hard for most westerners to understand the place of durian in South-East Asian societies – yet perhaps champagne plays a similar role in the West."

The early Western accounts of the durian went into raptures over the "King of Fruits," as it is called in Asia. Linschoot, writing in 1599, (quoted in Alan Davidson 1999) says: "It is of such an excellent taste that it surpasses in flavour all the other fruits of the world." The great naturalist Alfred Russel Wallace said,

> to eat Durian is a new sensation, worth a visit to the East to experience ... the more you eat of it, the less you feel inclined to stop.

> . . .

> Then there is a rich glutinous smoothness in the pulp which nothing else possesses, but which adds to its delicacy. It is neither acid, nor sweet, nor juicy, yet one feels the want of none of these qualities, for it is perfect as it is. (Quoted in Alan Davidson 1999: 263)

While these descriptions might prepare the reader for a delicious and unique fruit, you would never try even a mouthful if you read virtually any colonial or modern writers, who revile the durian and liken it to putrid garbage. Consider this:

> Among the charitable, printable comparisons: overripe cheese. Rotting fish. Unwashed socks. A city dump on a hot summers day. Historians report that Sir Stamford Raffles, who established Singapore as a British trading post in 1819, held his nose and ran in the other direction if he caught even a whiff of the dreaded fruit. Another former British governor likened the stench to carrion in custard. And yet the three million people of this prosperous island city-state cannot get enough of it. (Philip Shenon, New York Times, July 18, 1994)

I myself had an exchange on New York City radio in 2002 with Robert Sietsma, a popular writer of ethnic food guides. The show was about “undiscovered” ethnic foods. I suggested the durian. “Sure,” he said, “uh, I really like it – even though it tastes like vomit.”

The New York Times’ R.W. Apple Jr. is an excellent and sensitive food writer. Yet he can write, at the end of an article about how delicious he finds the mangosteen, another Asian fruit:
In this balance between yin and yang, mangosteens supply the cool element to offset the heat of the other most-loved Southeast Asian fruit, the huge, spiky durian, whose foul aroma would stun a goat. Many Asians therefore like to consume the two fruits at the same time.

“We describe the mangosteen as the queen of fruits,” he [a local food expert] reminded me. “We call the durian the king.”

Well, I for one have always preferred the company of ladies. (New York Times, September 24, 2003)

The reader might find it odd that Southeast Asians adore a fruit that “tastes like vomit” and smells like a sewer, and especially wonder how Westerners used to think it was deliciously “perfect as it is.” Why did Western meanings associated with the durian change? Conjecture (and analogy with Africa) suggests the development over time of distancing attitudes toward colonized peoples. The Africanist Basil Davidson (1995) points to such a shift in Europe's stance toward Africa. In ancient times, Davidson claims, Europe looked upon Africans as equals or better, as people “to learn from”; yet in preparation for the slave trade, in which Africans were treated as commodities, the European culture evolved a concept of the Africans as less than people – for how could you treat people the way the slave trade necessitated Africans be mistreated?

At the early time of Linschoot, Southeast Asians were viewed by Europeans as exotic, fascinating peoples from whom one could learn many things. As time went on, colonial interests became paramount, and it became functional for Western imperial cultures to create a representation of the peoples they were trying to conquer and exploit as less than human, less worthy of rights, who only deserved to serve the colonial apparatuses. We can conjecture that perhaps this distancing and devaluing fed the shift from the Western reality of the perfect durian to the Western reality of the foul, putrid durian. In any event, while the durian did not change, its meaning certainly did. Food, conceivably the substance most assumed to have intrinsic meaning, has no meaning except what is constructed through conceptual systems. I am of course not suggesting that Westerners who find the durian repulsive are harboring anti-Asian feelings; I am claiming they have learned an arbitrary conceptual code system in which the durian is devalued as putrid and inedible.

3. The Swahili and Lamu town

3.1 Swahili society and the path to heshima

Knowledge of a code can identify us as members of our group, and performance of this code validates our claim to a specific position in the group. Through a confu-
ence of the nonlinguistic meaning systems of food, the construction of social and gender identity, and the use of architectural, private, and urban space (and possibly other systems), the validity of this claim is articulated in the ancient Swahili city-state of Lamu, Kenya.

Fieldwork in Lamu reveals a highly coherent set of specific learned behaviors that actors can choose to perform in certain contexts to signal meanings on a continuum. These meanings span a semantic substance I hypothesize to be the “validity of one’s claim to position of respect in the group,” an important organizing principle in Swahili society.

![Figure 1. Heshima meanings](image)

The desired end-result of behavior is conceptualized as *heshima*, one of the attainments most sought after in Swahili society. I translate *heshima* as ‘respect, dignity, honor’ including ‘the rules and rights to dignity.’ One who has *heshima* has prestige. At the other extreme is *aibu* ‘shame,’ which can be seen as the absence of *heshima*. Advice I have often heard given is *usivunje heshima, hutaki aibu*, ‘do not lose [lit. break] your *heshima*; you do not want *aibu*.

As the anthropologist Ntarangwi explains,

> When a Swahili person engages in an act that contravenes the expected measure of *heshima*, they are said to have brought shame and disgrace (*aibu*) not only to themselves as individuals but to their families and entire community. These notions of *heshima* and *aibu* intersect with Swahili social stratification … evidence of this social stratification often finds expression in people’s public conduct and interaction… A Swahili man or woman of noble ancestry (*muungwana*) will be more concerned about *heshima* than will a Swahili of slave ancestry (*mzalia*). The *muungwana* [in Lamu kijoho] will be careful to select his/her language, places to visit, who to interact with, and the mode of dress, since all these contribute to the projection and maintenance of *heshima*. A woman from a high-status family will be more restricted in the house than a woman from a low-status family … because the public space is regarded as male space and for a woman to occupy such space is to lower her social status. (Ntarangwi 2003: 120–121)
Ntarangwi’s analysis is quite congruent with our original analysis (Leonard and Saliba 1992), although the two use slightly different data sets. Like us, he notes that essential for *heshima* are behaviors, not only birth: the rules of language, the places one visits, interactions with people, and dress. Our data in this paper consist of overt signaling behaviors not including language but including the places one visits in the city of Lamu, interactions with people, dress (tangentially), and, not mentioned in Ntarangwi’s discussion, food behavior and use of architectural space. Further, Ntarangwi and we see the same intersection of social class with what we call avoidance of intimacy: the more upper class, the more avoidance of intimacy.

However, while I would argue that Ntarangwi’s statement above is correct, I would add what I believe is a crucial distinction: being concerned with *heshima* is different than being able to signal higher levels of *heshima*. Ntarangwi states that nobles will be “more concerned about *heshima*.” Yes, but the noble has access to signals on a higher portion of the *heshima* scale than does the non-noble, regardless of how concerned a non-noble – including those of very low-status ancestry – may be. Indeed the advice I quoted above, warning to protect one’s *heshima* – *visi-vunje heshima, hutaki aibu* – I heard most commonly from lower status speakers.

Hirsch (1998), a linguistic anthropologist studying Kenyan Islamic courts, states: “Swahili people understand *heshima* as represented by an individual’s charity, good descent, propriety in behavior, and cleanliness” (Hirsch 1998: 298 fn.). It is a “reputation, secured through behavior that is not shameful (*aibu*)” (Hirsch 1998: 48).

An individual has considerable control over his or her *heshima*, and, as an ideal, all interactions should be undertaken with the intention that they will reflect and augment one’s own *heshima* and that of others. For women, *heshima* is more directly linked to sexuality and to “uncivilized” (*ushenzi*) behavior than for men. At the same time as women are ordered not to “break *heshima*” (*vunja heshima*) by behaving in ways that cast doubt on their sexual purity, men are urged to have *heshima* for women by refraining from actions that would violate their purity. A woman’s *heshima* is called into question by any avoidable contact with a man to whom she is not related as *maharimu*, a category of persons within certain degrees of consanguinity and so forbidden to marry or have sex with each other. Men secure their claims to *heshima* as honor through the *heshima* of the women of their households. (Hirsch 1998: 48–49)

A lower-class person in Swahili society aspires to *heshima* appropriate for his or her position and cannot imagine attaining the level of *heshima* an upper-class person will, because the upper-class *kijoho* starts out with more noble descent, and, we may add, many more financial and physical resources to maintain strict propriety (e.g., a noble wife can stay secluded and not need to work outside the home). The lower-class person seeks to maximize his or her level, given the reality of resources and position in society.
This has a linguistic analog. Labov’s well-known work on variation theory (e.g., Labov 1972) demonstrates how certain linguistic features vary systematically according to social factors such as gender, class, and formality register. The speech of higher socio-economic classes more often correlates with overtly high prestige sociolinguistic variables, for example postvocalic \( r \) in New York City. In New York City, an upper-middle class speaker is more likely than a lower-class speaker to pronounce postvocalic \( r \). The more attention any New York speaker pays to his or her speech (because of, say, the high formality of a speech situation), a higher percentage of \( r \) will they utter in words like car and guard. New York City speakers all know that moving toward more \( r \) is prestigious. But the high-attention end of a lower-class speaker is not as high as the high-attention end of an upper-middle-class speaker, and the lower end is lower. In other words, even with the greatest of desires to perform at the high end of the scale, the repertoire of behaviors available to individuals is dependent on their preexisting (that is, preexisting the utterance) socio-economic identities (because their idiolects were developed in those particular milieus).

The current analysis posits variation in nonlinguistic meaningful features according to similar social factors. Thus we add, in Figure 2, to our scalar schematic some types of actors and types of behaviors:

<table>
<thead>
<tr>
<th>Validity of Claim of Position</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH (HESHIMA ‘RESPECT’)</td>
<td>LOW (AIBU ‘SHAME’)</td>
</tr>
<tr>
<td>Upper classes (especially women)</td>
<td>Lower classes</td>
</tr>
<tr>
<td>more restriction of interaction</td>
<td>less restriction</td>
</tr>
<tr>
<td>more privacy</td>
<td>less privacy</td>
</tr>
<tr>
<td>more avoidance of intimacy</td>
<td>less avoidance of intimacy</td>
</tr>
</tbody>
</table>

**Figure 2.** Heshima meanings, with actors and behaviors

3.2 Food and drink behavior along the public-private continuum

Spatial orientation, like language, comes into being through social contract. As Dolgin et al. (1977) write:
Whatever the natural aspects of space, it is, like every other part of nature, given meaningful configuration in the culture of any people, and different peoples construct space differently, to a greater or lesser degree. (Dolgin et al. 1977: 131. See also Bourdieu 1971 and Donley-Reid 1990.)

The cases of garlic and of the durian recounted above have shown how valutative meaning-systems organize and mediate human experience. Examples of Swahili place give similar evidence. Analysis of Swahili private architectural space – the traditional Lamu Swahili stone house – has provided insight into the organization and divisions of Swahili culture. Allen (1972, 1979) and Ghaidan (1975: 75) analyze the house in terms of an “intimacy gradient” that explains both the architecture and the identity of persons allowed within each space. Middleton (1992: 66) speaks of its “several degrees of privacy and purity.”

These analyses of Swahili private space show the meaningfulness of spatial divisions in the Swahili stone house. The outer skin of the house consists of high blank walls. Non-family members normally remain in an entrance porch outside the house. Inside there are rooms around an open courtyard. The first space inside is a separated guest area. Each area within allows only closer and closer family members to enter. These areas are not necessarily separated by walls, and even walls often have large open doorways. Each area is marked off by being several inches higher than the level before. Allen (personal communication, 1976) explains how, in a manner I find reminiscent of language, the Swahili stone house has meaningful features that I found are below the level of normal consciousness, notably, the architectural feature of a small step up each time one passes through an invisible boundary into a more intimate area of the Swahili house. I discovered during discussions about the stone house with Swahilis who lived in such houses that they were often not aware that the different areas were marked by slight differences in height.

While Ghaidan (1975: 61) treats the “morphology of the Swahili town through an analysis of three related Swahili behavior systems” – pedigree, involvement, and privacy – he limits his discussion of privacy to the Swahili house.

Like the Egyptian temple, the Swahili house is axial in plan; as one follows the axis in the direction of the interior of the house, spaces become gradually darker, more decorated and more intimate. The axis acts as an intimacy gradient; the further up a room is placed, the more private it is. The gradient is marked at its ends by the daka [entrance porch] and ndani [the ‘inside’ room—parents’ bedroom suite], the most formal and informal spaces respectively; the former is the antithesis to privacy, the latter the antithesis to involvement. (Ghaidan 1975: 75–76)

We suggest, however, that the entrance porch is the “antithesis to privacy” only when considering the house apart from the rest of the world. If we add public spaces into the calculation, we find the porch is only perhaps midway on a more
comprehensive continuum of public versus private space. This public-private continuum begins in the house and extends outside, beyond the semi-public space of the entrance porch into truly public areas. Further, we observe, food and drink behaviors strongly vary depending on where one is along the continuum.

In Lamu, during the time of this research a city of some 10,000, society was rather traditional compared to other Kenyan urban areas such as Nairobi or Mombasa. Women and men in Swahili society are subject to purdah and women wear the buibui, a loose black garment with a veil, although it is generally held that purdah in Swahili society is less strict than in much of the Arab world (e.g., Knappert 1979: xix.). But among Swahilis, Lamu is by far the more conservative. Strobel (1979) notes:

[In] the past twenty years the custom of veiling has declined... the vast majority of Muslim women still wear the buibui, though most do not cover their faces and chests. Mombasa is in the vanguard in this respect... in Lamu, women in the streets show only their eyes. (Strobel 1979: 123)

3.3 The street scene

On Lamu’s streets we can observe groups of men on benches talking, drinking from small cups, and smoking; children carrying (and eating from) trays of deep-fried foods; men eating meals at outdoor tables while meat is charcoal-grilled nearby. A group of women enter a restaurant to eat snacks and drink fruit juice while passers-by purchase sweets from a shop to take home. A youth buys a bundle of leafy twigs from a man sitting at the roadside. A first look at the participants here might not reveal how highly structured are the cultural rules governing food behaviors, and how different they are from what outsiders like ourselves might initially expect.

To New Yorkers, for example, it is quite natural for men and women of different socio-economic classes to stroll about the street eating and drinking a broad variety of food and beverages (although there are constraints, to be sure). But in Swahili culture, only small children are suffered to actually eat while walking on the street (Maryam Abdallah Yahya interview, June 1990). There is even a saying that warns against such behavior, Kula kitu kitamu kinjianjia mtu atasibiwa na shetani ‘One who eats delicious food while walking will be possessed by a devil.’ The annotator of a book on Swahili sayings explains this to mean, “It is regarded as bad manners to eat while walking” (Farsi 1958: 22).

The proscription against eating while walking extends also to eating, or even drinking, while standing on the street. One Mombasan said, “Even to drink water one must be seated [if] on the street” (Ahmed Samatar interview, June 1990). Men can drink on the street if seated. Swahili sellers of kahawa chungu or kahawa tamu (coffee and cinnamon beverages) serve customers who sit on benches at the side
of the street. Other vendors serve *maji tamu* ‘fruit juice.’ Men may eat outdoors if seated in a restaurant (called *hoteli* in Swahili). Such outdoor seating is normally separated from the street, if only symbolically, by a low fence.

Women do not patronize *kahawa*-sellers at all, and at restaurants they do not sit at outside tables. If women eat at a restaurant – a fairly uncommon occurrence – it is often in an all-female group, in an inner, enclosed room away from the other patrons. One sometimes sees exceptions in restaurants that cater to travelers. Indeed travelers are an exception to many rules, for example, fasting during the month of Ramadhan. A Swahili proverb says *Msafiri kafiri,* ‘A traveler is (lowered to the level of) an unbeliever’; that is, one must do all sorts of things in the course of traveling that one would normally not do, e.g., eat unclean foods and break other taboos. Another related proverb says *Msafiri ajapokuwa ni mfalme ni maskini,* ‘A traveler, even though he be a king, is a poor man.’ (Versions of this and other proverbs quoted are found in Farsi 1958.)

Young girls are not subject to purdah, and children of both sexes travel the streets to restaurants to deliver food that has been prepared by women in their families. Often extra food is prepared and the children sell it on the street to other children, who can consume it then and there, or to adults, who will eat at home (Maryam Abdallah Yahya interview, 1991).

To summarize, while children are fairly free even while walking, it is not proper for adults to drink or eat while walking or standing in public. Men can drink *kahawa* seated at the roadside and can eat food at outdoor tables of restaurants, though some barrier often suggests the division from the street. It is proper for women to eat in the inner rooms of restaurants. Ntarangwi points out (personal communication, 2005) that this is because Swahili public space is considered male space, and thus women can only occupy it when they are shielded from the male gaze by wearing a *buibui* or being in an inner room away from the public eye.

### 3.4 The upper class and travelers

The upper class, however, most notably wealthy noble-borns, avoid even restaurants. Ghaidan (1975) writes concerning his 1969–1973 fieldwork in Lamu:

> All the town’s cafes are concentrated in Langoni [the part of town where shops and lower-class residences are located]. A *Kijoho* [descendant of an influential lineage, referred to in Mombasa as *Muungwana* as in the Ntarangwi passage above] would not normally be seen in one. When the present writer, out of ignorance, insisted on one joining him for coffee, the cafe was soon surrounded by an astonished crowd. It was then explained that that was the first time that a *Kijoho* had ever sat in a cafe. Usually, house entrance porches [*daka*, mentioned in the Ghaidan quotation above] were used for this purpose. Visitors normally arrived after supper to
spend the evening talking, fingering beads, and eating halwa [Swahili: halua] with coffee. Beads, coffee and halwa are traditional adjuncts of talk in Swahili society; one of the captains in Vasco da Gama’s fleet was treated to halwa by the Sheikh of Mozambique in 1498. He was also made a present of a string of black pearl beads. (Ghaidan 1975: 65)

The daka entrance porch of a stone house, where Ghaidan says he should have had coffee (and halua) with the kijoho, occupies an interesting position. It conjoins the private and the public. Lamu streets are narrow, sometimes less in width than one’s outstretched arms, and bounded by high blank walls that enclose houses and their internal courtyards. The entrance porch, lined with stone seats, is two or three steps up from the street and built as a deep recess in the wall. It is in view of passersby. At its end is a door leading to the house. The porch is neither home nor street. If Ghaidan is accurate, it is the most public setting in which the upper class cares to eat or drink. (We can safely predict that someone who will not frequent a restaurant will certainly not patronize a sidewalk kahawa seller, a place even further towards the public end of the continuum.)

Since halua, as Ghaidan notes, is the traditional accompaniment to kahawa, we might expect to find it for sale by sidewalk kahawa-sellers. But halua cannot be bought on the street. One buys it at halua-shops to take home or to eat inside the shop (with kahawa). Some streetside kahawa sellers offer only kahawa, others sell betel-leaf, cigarettes, and often sugar-candies. Although halua can be eaten on a daka porch, it apparently is too much a “food” to be eaten at streetside. It was pointed out to us by a Mombasan that halua in Mombasa used to be, and in Lamu still is, sold sewn up in very small baskets. “This makes it clear that this is not something to eat in the street, but that you will wait to get home to eat” (Ahmed Samatar interview, June 1992).

For Swahilis, meals are normally taken in the home. Restaurants are typically for those who are traveling, or working far from where they live. Foreign visitors sometimes puzzle over why Swahili restaurants do not attempt to serve food in the fashion it would be served in a Swahili home. The reason, we believe, is that the intent of a restaurant is not to try to recreate the privacy and intimacy of home. This would be rather difficult, given the meanings of space and food interaction in a Swahili town. Ntarangwi agrees, and cogently points out (personal communication, 2005) that eating at home also fulfils the collective logic espoused in much of Swahili culture where people not only eat together but also often eat from the same sinia, a large food tray. In restaurants, each person is served individually. Further, he notes, home-cooked food is considered superior to restaurant food because of the former’s identification with the preparer (sister, mother, wife, grandmother, etc.) and thus carries the social relation between the person who prepared it and the consumer.
4. Analysis

If we look at food and drink behavior patterns we see that relevant features concern spatial occupancy (where one is), spatial activity (standing, walking, sitting), and type of food. Also relevant are age, sex, and rank. Together these intersect in a matrix of possible behaviors, the typically appropriate intersection seen as capable of maintaining *heshima*. Figure 3 summarizes the elements. Contrasting features, as diagrammed in Figure 3, include walking versus standing, standing versus sitting, the street versus other public areas, outer seating versus inner enclosed seating, children versus adults, men versus women, lower class versus important lineages.

Foods and other consumables are roughly organized into drink, non-food, and food proper. Drink includes water, *kahawa*, fruit juice and *dafu* ‘unripe coconut juice (and sometimes its gelatinous meat).’ Non-food includes *miraa* and the sugar candy or gum chewed with it to counteract its bitterness, cigarettes, betel and the sugar candies taken with *kahawa*. Food has its own continuum from snacks to home meals. Restaurant meals are typically not eaten by important lineages; snacks like halua are taken by all.

These features seem to combine in the following way: food and drink activity is most curtailed at the extreme public end of the Spatial occupancy / Architecture continuum, *walking* (and, usually, *standing on the street*). Here, with quite minor exceptions, no one is allowed to eat or drink save children, to whom mature societal rules do not yet apply. One degree along the continuum, *sitting on the street*, drink and nonfood are allowed, but only for men. At a further degree, *sitting somewhat separated from the street*, (outdoor seats in a restaurant) food proper is allowed, but only for men. Another degree along, *sitting outside the home enclosed by walls*, i.e., an inner room at a restaurant, women are allowed to drink and eat. One more degree, *inside the home*, all foods are allowed (and space and activity is subject to further, intimacy gradient rules). The upper class do not consume past the *daka* entrance porch (whose position on the continuum is towards the privacy side by being in the upper-class, stone-house part of town).
Spatial Activity
Left: wider range of foods allowed, wider range of places, by wider range of identities; higher conservation of heshima,
Right: fewer allowable foods, more danger of abu

| sitting | standing | walking |

Spatial Occupancy/Architecture
Left: more private, more identities are allowed to eat more types of food, higher conservation of heshima,
Right: more public, fewer allowable foods, more danger of abu

house: ‘inside’ room
  inner living room
  outer living room
  guest areas
  (entrance porch)
  restaurant: enclosed seating
  outer seating
  street: seating
  street itself

Identity
For each subheading, left has more constraints, easier to lose heshima

| age: | adults | children |
| sex: | women | men |
| rank: | important lineages (i.e., vijo, nobles) | lower classes |

Section of Lamu
Left has more privacy, more inherent heshima

Mkomani: Upper class stone old town
Langoni: Commercial and lower class residences

Type of Food and Other Consumables Allowed
Left: more constraints on where and by whom allowed
Right: fewer constraints

| home meals | snacks/ restaurant meals | non-food | drinks |

Figure 3. Matrices of food and drink behavior: Spatial activity, spatial occupancy, age, sex, rank and category of food
While the public-private continuum operates as a *heshima* control for the whole society, the lower and upper classes range along different parts of the continuum; the upper stay on the extreme private end and the *daka* porch is the far point of their public food and drink behavior. Limiting one’s operations at the higher end of the scale will, all things being equal, result in greater *heshima*.

*Heshima* is involved with intimacy, as is the very intimacy gradient at work within the Swahili house. At all levels of social stratification in Lamu one does not want forced intimacy. Control of intimacy means greater control of *heshima* and *aibu*. Those with the means to control intimacy do so. A purdah society like that of the Swahili is characterized by ritual segregation of women and men and the protection of the bodies of women from the sight of men. Eating and drinking for women therefore carries the risk of forced intimacy over and above that experienced by men because of the possible necessity to reveal parts of their bodies in the act of ingesting. In such circumstances the purdah garment itself can be used to conceal eating and drinking – women will convey food to their mouths under their veils, or even have the veil shroud the food plate itself – and thus the garment has an analogous function to the house in providing eating privacy.

5. Discussion

5.1 If eating is inherently intimate, how can food be arbitrary?

Even though certain meaningful signals have non-arbitrary origins, this does not exclude those signal-meaning pairs from being arbitrary for learners of the code. Signals are not learned with histories attached. The tomato originated in the new world, yet is now authentically Italian. Compare American Sign Language. Many signs have non-arbitrary origins yet are as arbitrary as *cat* – or, shall I say, *meow*. Cats do not make the noise made by English speakers saying *meow*. Even what we may call *motivated* signs, like *meow*, or many ASL signs, are still arbitrary.

This Swahili complex of signaling behaviors involves arbitrary meanings: witness the inability of untrained individuals to negotiate the appropriate behaviors. Outsiders, Euro-American or from other parts of Kenya, fail again and again to avoid *aibu*-producing behaviors until schooled by Lamuan locals, and often well after that.

5.2 Ramadhan proofs the rule

Food plays a central role in the important Muslim month of Ramadhan. Food behavior comes under intense scrutiny, because all devout Muslims fast during daylight hours. They are forbidden to eat food, chew *miraa*, smoke cigarettes, drink
water, or, I was told many times, even swallow their saliva (and there are endless disputes about this). The evening call to prayer signals each day that the fast is over. Women break the fast at home, as would be predicted by the _heshima_ analysis. Many men, however, break the fast in the street, contrary to what we might expect from the _heshima_ analysis. In preparation for breaking the fast, table upon table of foods including snacks and restaurant foods, sweets and savories are set up right in the street in Lamu and Mombasa. _Kahawa_ ‘coffee’ and _majì tamu_ ‘fruit juice’ are also set ready. Men pay for their food and drink and, in Lamu, lay it on tables, sit opposite it and wait for the call to prayer that will signal the end of the day. In Mombasa men sit or stand in groups on the street and pool the food they will break fast with.\textsuperscript{13} After eating, often standing, to break the fast, the men quickly go off to the mosque to pray, and then home to eat a full meal.

Why are men apparently breaking the rules of polite eating conduct? That is, why are they signaling apparently _heshima_-breaking behavior meanings by eating food and drinking beverages standing or sitting on the street surrounded by strangers? These actions are inappropriate in terms of publicness, high intimacy, high involvement, and assertion of sameness as oneself.

We have previously (Leonard and Saliba 1992) analyzed Ramadhan fasting in Swahili communities as a _public group act_ testifying to the solidarity of an Islamic community. In light of this, the meanings signaled by eating on the street, inappropriate in a normal context, are highly appropriate in the special context of a religious ritual whose very essence is the _unity_ and _solidarity of the community_. Actors have choice. Many times, in usual contexts, people will perform an action other than that which is most expected of them, given who they are. They may signal meanings on the higher end of the scale and run counter to fellow members’ predictions. This may be received well, or received as inappropriate behavior that will not therefore accrue _heshima_. Often actors signal meanings lower on the scale than expected to, and lose _heshima_.

In the Ramadhan situation, although the actors signal meanings far lower than normal expectations, they intend the inference to be made that in this context it is more _heshima_-worthy to signal meanings of intimacy in a quite public, all-inclusive setting. In this context the larger goals of the society are inferred to take precedence and therefore these normally _heshima_-breaking behaviors instead garner _heshima_.

Thus, what here constitute _heshima_-worthy behaviors are shifted along the scale. Just as speakers signal linguistic meanings that contribute to the overall intended message depending on the context, actors here signal meanings in a particular context to convey a message. The intended inference here is that during Ramadhan one wants to signal publicness of fasting and following religious guidance. During the context of Ramadhan, all men are as in the same Islamic family.
and the very town is as their Islamic house. Behaviors of high involvement and intimacy here serve to establish *heshima* rather than *aibu*.

6. Conclusion

6.1 Meaning in anthropology

There is a vast anthropological literature on the meaning of nonlinguistic behaviors. Levi-Strauss, for instance, well-known for his structuralist treatments, deals with meanings attached to a wide array of nonlinguistic phenomena, notably food, e.g., his famous opposition of the raw to the cooked. I would argue that for Levi-Strauss, sufficient explanation of the data consists of the categorization of that data into dyads or triads. This relates directly to his orientation, his construct of the collective unconsciousness of *l'esprit humain* or the human mind. Levi-Strauss’s foremost English interpreter Edmund Leach (1974) explains that Levi-Strauss claims the human brain embodies “certain capacities for making +/- distinctions, for treating the binary pairs thus formed as related couples, and for manipulating these ‘relations’ as a matrix algebra.” How does Levi-Strauss know that the human mind can embody certain +/- distinctions? Leach (1974: 52–53) attributes this knowledge to structuralist linguistics, specifically Jakobsonian distinctive features.

Levi-Strauss has been criticized by other anthropologists on a number of levels, including his *a priori* framework of +/- distinctions and his assumption of an autonomous Culture paradigm. Much current theoretical debate in anthropology revolves around the viability of the Culture paradigm and its ability to explain behavior. Some theorists claim that autonomous “cultural systems” are not the adequate explanatory devices they are claimed to be. Whitehead (2000), introducing the theoretical ramifications of her study of food sharing and taboo in Papua New Guinea, takes exception to the autonomous Culture paradigm:

> Commonality and ease of acquisition does not, in the view of the strict Cultural Anthropologist, guarantee that [a] feature is “human nature.” The most glamorous arguments in the subdiscipline are typically those aimed at rolling back any human nature part of the equation even further than the previous generation of scholars has managed to do... Culture is active and causative; it is the inscriber. The mind is passive and receptive; it is the written upon. As so much of what is interesting in this duo occurs on the side of culture, it follows that culture itself is the source of much that appears “system-like” in human life-ways. (Whitehead 2000: 3–4)

Whitehead continues that the extreme form of the “culture” view does not need to explain further than to posit that some behavior is culturally determined, as other animal behaviors are genetically determined. She counters with examples of food
taboos that other analysts saw as arbitrary and solely the product of autonomous culture but, she found, upon analysis were quite practically motivated, e.g., her discovery that many tabooed food items were simply too difficult, small, etc., or otherwise counterproductive to gather.

I would stress, however, that for learners of the system, unless they have had explained to them the histories of the difficulties in obtaining a certain item, and that that often results in a taboo on that item, the taboo value on that item is learned in as arbitrary a fashion as English speakers learn that *cat* means ‘cat’.

Whitehead herself does not claim autonomous culture and arbitrariness are one and the same thing. Discussing signifiers and signifieds (personal communication, 2005), she supports the arbitrary association of signal and meaning:

> The [Papua New Guinea] Seltaman distributional taboos … are a sort of collective stamp of approval placed on the dynamical outcome of repetitive decision-making in the village sharing system. You could never penetrate them without knowing the logic of those repetitive decisions and this logic has nothing to do with signifier/signified similarities. It certainly involves cognitive operations, however. (Whitehead, personal communication, 2005)

This overall model is similar to what I am suggesting for the behaviors that signal meanings associated with *heshima*. The meaning of a signal, linguistic or nonlinguistic, is culturally assigned because the users of the language or other system want to use that meaning in their system and because the signal is available, not because a preordained symmetry requires it.

6.2 Meaning in Columbia School linguistics

If, as Diver (1975: 12–13) and Columbia School in general maintains, “language is not an activity that is unique within the species” but is instead an instance of human behavior, then it should not be surprising that other instances of human behavior are amenable to analysis along lines similar to CS linguistic analysis and in terms of CS constructs. I will conclude by making explicit the similarities and differences I see between the behavioral analysis presented in this paper (including possible implications of it) and CS linguistic analysis, including the overall theoretical approach. This demonstration will consist of quotations and paraphrases of passages from two foundational CS documents, Diver (1975) and Diver (1995), with the original words that I have changed given in double quotation marks in square brackets.

First the similarities:

> Theory is extrapolated from analysis, as a generalization over the results of individual analyses, rather than being set up in advance as a guide to analysis.
The theory spells out the way in which the problem is solved. The problem is posed by certain physical, visual (“acoustic”), observations – the movements (“sound waves”) produced when a person acts (“talks”).

The problem is solved by establishing what motivated the actor (“speaker”) to produce those particular movements (“sound waves”). What began as a problem in physics is thereby transformed to a problem in human behavior...

The general motivations (here termed ‘orientations’) discovered by analysis are derived... from normal traits of human behavior... and from principles of communication (Diver 1995: 110).

It is of course no novelty to associate social behavior (“language”) with communication. The critical point, however, is not whether such behavior (“language”) is used for communication (or the extent to which it is used for purposes other than communication) but whether its very design and structure are directly motivated by the act of communication. (Diver 1975: 3)

The social behavioral (“grammatical”) units are organized into systems in which a semantic substance is completely divided up into parts, and each part assigned a signal. Each part is called a meaning...

The internal organization of the system [involves] relations of “value” ... (Diver 1995: 112).

The semantic substances themselves recur to a very considerable extent from culture to culture (“language to language”) – although they are by no means universal – but the manner of dividing the substance up differs. At the same time, to come back to the human factor, the recurrent character of semantic substances ... relate[s] in a very natural way to the pervasive interests of humans: What are you doing where? Who are you? When are you doing it? (“What happened? Who did it? When? Did it really happen? Who else was involved? Don’t talk so fast, will you, I can’t understand you.”)

The recurrent semantic substances, varying constantly in detail from language to language, address themselves to problems of this kind. Can it be doubted that the communicative factor is playing a role in the structure of language? These are not the interests of the logician and philosopher, who, indeed, have had to devise “languages” of their own to assist them with the problems involved in an accurate analysis of truth value. They are rather the problems of the man in the street, and all the research we have done along these lines indicates quite clearly that it is his communicative interests that have shaped his communicative instrument, and this again, regardless of how little the individual may know of exactly what he is doing. (Diver 1975: 19–20)

The overall network of relationships may be viewed as having communication as its driving force, its principal motivation...

The hypotheses themselves cannot be said to have been “learned” by the actor (“speaker”). Rather, they have been devised on the basis of observation of the performance of others, particularly the relation between what people do (“say”) and the social context in which they do (“say”) it...
These hypotheses of course cannot be observed directly. They are inferred by analysts on the basis of the way people behave [“communicate”]. When we study the behavior [“process of communication”] in great detail, we come to the decision that speakers are behaving as though they are operating with a particular network of particular hypotheses. Coming to that decision is what the task of the analyst is all about (Diver 1995: 113).

The differences, too, are considerable between linguistics and the problems considered in this paper. The physical nature of the phenomena being observed is that of the movement of the whole human body and of foods, not that of “sound waves produced when a person talks” (Diver 1995: 110). The constraints on the behavior are consequently grossly different, as they do not involve “the control of the musculature of the vocal tract” (Diver 1995: 62). The medium of observation is visual, not “acoustic” (Diver 1995: 48). The signals of the meanings involve behaviors with objects in space: posture and movement vis-à-vis various foods within commercial, public, and domestic architecture; not “morphemes,” “word order,” and significant absence of morphemes or “zero” (Diver 1995: 84–85). All these differences are directly related to the nature of the problem chosen for analysis. More telling, perhaps, is an apparent difference involving the meanings proposed here. There so far does not appear to be an important difference between the signaled meanings and the messages communicated by and inferred from those meanings; that is, between the “hints offered by” the actor and the “guess” made by a casual observer. Specifically, we have not here made a crucial distinction between respect or shame and any further message conveyed or drawn from those meanings in conjunction with other, contextual meanings.15

All this suggests that while the use of food and architecture in Swahili culture may be less “versatile” (Diver 1975: 9) in communication than is human language, it is similar in some interesting ways to human language and is amenable to Columbia School-inspired analysis.

Acknowledgements

Many thanks to Professor William Diver, who encouraged me to analyze behaviors beyond language and who guided me through the early stages of this analysis. I am grateful to the editors of this volume for their support and hard work; Joseph Davis deserves special thanks for suggesting ideas for the concluding comparisons with CS theory. Thanks go to Ricardo Otheguy and Wallis Reid, whose penetrating criticisms of earlier versions of my analysis enabled me to finally see what the real issues are. After I had finished a late draft of this paper I happened by good fortune to meet Mwenda Ntarangwi and he made several valuable comments as well. Ben-
ji Wald's profound insights regarding linguistic theory and Swahili language and culture have helped guide me throughout. Thanks also to an anonymous reviewer who gave many useful criticisms and suggestions. I am, of course, responsible for any mistakes that made it through to the final version.

Notes


2. Two specialties within linguistics contribute to the analysis in this paper. One, theoretical semantics, seeks to understand how language creates and communicates meaning. The second, sociolinguistics, focuses on the interplay between language and social variables, for example, group membership, class and gender.

   I was led to the concerns of this paper – sign-systems in a context larger than language – by what I suppose is a somewhat atypical research background for a theoretical linguist. At Columbia I was influenced by two professors who had quite different theoretical concerns and perspectives on linguistic research: William Diver, founder of the Columbia School, and the sociolinguist William Labov.

   Diver trained his students in Saussurean sign theory and textual analysis. Two opposing traditions within linguistics attempt either (1) to analyze language through a priori mathematical and logical categories, as do Chomskyian schools of thought; or (2) to demonstrate, as did Diver, that language is an instance of a human communicative system and that the design of language is shaped by that. This second tradition, exemplified by the Columbia School of linguistic analysis that Diver founded, has produced a number of successful analyses of natural languages, especially over the last forty or so years. Columbia School theory holds that language consists of signal-meaning pairs, that is, Saussure's signs, and that signs are organized into systems where meanings are defined in opposition to other meanings. Thus the theory takes Saussurean insights quite seriously (while departing from Saussure in some matters), and over the last few decades its investigators have rigorously tested the theory against naturally occurring speech in dozens of languages, achieving progress in understanding the nature of meaning and linguistic communication. (See, for example, Contini-Morava and Goldberg 1995; Reid, Otheguy, and Stern 2002; Contini-Morava, Kirsner, and Rodríguez-Bachiller 2004.)

   William Labov prepared his students for variation theory fieldwork, and for me this meant fieldwork in Africa and Thailand. An important aspect of Labovian training was “entering the speech community,” i.e., learning how to gain acceptance and conduct fieldwork that would produce naturally occurring vernacular ingroup speech even in the most foreign of societies. This required a detailed knowledge of the culture, and social categories within the speech community.
3. While the use of the term *meaning* varies enormously in linguistics and anthropology, Diver (1995: 72–77, and earlier) largely confines the term meaning to the semantic content of grammatical signals, less of lexical items and certainly not of social behaviors. I am explicitly attempting to extend the Diverian concept into the last.

4. I followed this (Leonard 1995a) with a semiological study of an alcoholic beverage I had been taught to brew and ceremonially use by the elders of the Akamba people – since I was occasionally drafted to sit on their Elders’ Council – when I lived in their district in the 1980’s. This study defines beverages by their value rather than substance but is a much simpler type of analysis than that of the street/public foods.

5. For 15 years I have sent teams of students out to research food behavior and attitudes among both American and non-American groups, using food categories Saliba and I developed in the 1980s and published in 1999a and 1999b. We define for example, insider food: “group members feel that to truly be a member of the group, one must eat this food. Further, only they eat this food; outsiders do not eat it because they are unwilling or physically unable to eat it, or because it is unknown to them.” (1999a: 174)


7. Jews, said the Nazi terrorists, were using their radios to contact the English air force and were directing the bombing of Hamburg. Furthermore Jews had stolen so much money from true Germans that they all had silverware illegitimately.

   We went home, picked up our radio and our silverware and the whole family walked to the nearest precinct house as we had been ordered. A large crowd of Jews was already there. As each Jew lined up to enter the police station his name was checked off on a list prepared there. “It stinks from garlic here,” shouted a policeman while others enjoyed a number of offensive “jokes” at the expense of the Jews.

   It was the last Yom Kippur for the Jews of Hamburg. Deported, murdered and scattered to ‘the four corners of the earth,’ only a few survived. (Falk)

   We learned that Christ was God’s son, not the son of some Hebrew. Besides that, the Catholic Church teaches that Holy Mary, the mother of Christ, died a virgin. That means she never had sexual relations with a man, certainly not with a flat-footed garlic Jew. (Springer 1939)

   Also many other stories, including the one of a group overnight ceasing to use garlic, in excerpts from Neiman (1992).


9. It is certainly true that the durian has an abundance of sulfur-related constituents and when overripe or rotten the fruit smells putrid in the way only sulfur compounds can – certainly worse than many other decomposing fruits. But this is true of many foods. Does fresh fish smell bad? Or fresh meat? An anonymous reviewer of this paper noted:

   A relevant comparison would be strong cheese, which is considered a great delicacy by some in the West despite its putrid odor. In fact, the valuing of putridity is arguably the same sort of arbitrariness as that associated with garlic in the previous discussion, although it can also be argued that since putrid things are more likely to be diseased or indigestible, there is a non-arbitrary reason for classifying putrid as ‘non-food’. It is worth noting that
Southeast Asians themselves are aware of the nature of durian's smell and treat it exactly as the French treat smelly cheese. Despite many insightful suggestions elsewhere, the reviewer here is wrong, and repeats a common misperception of durian. Durian is nothing like French smelly cheese; durian is not prized as a fermented food (a very highly valued Southeast Asian food group, to be sure), and more to the point, durian has no putrid smell unless high or spoiled. Durian is banned on airlines and in many hotels, if as a practical matter, because once it goes bad, and it can go bad quickly, the sulfur compounds penetrate and linger. But we are here analyzing the durian as food, not in its role as old garbage. Whereas virtually all the modern writings mention putridity, nowhere in the early Western writings do we find mention of a putrid smell; the closest we find to any displeasure is Wallace (1869 [1999: 263]) saying he sometimes perceives incongruous notes like “onion sauce and brown sherry.” No doubt some part, at least, of Asians’ “admitting” durian smells bad is simply acknowledging modern Western opinions. Read between the lines of a New York Times account by Marian Burros:

Thais are often quite certain that Westerners will not like durian; many Thais don’t like it. Nat Boonthanakit, public-relations officer for the Tourism Authority of Thailand in New York, said that some members of his family “simply can’t stand the thought of having a durian in their house.” (New York Times, March 7, 1987)

But Burros finds differently:

I expected to be bowled over by the odor when I opened the durian I bought. What a surprise to find that it was not at all unpleasant and that the pale yellow flesh had an appealingly creamy, custardlike texture and a very sweet taste. My Thai friends, it seemed, were wrong. (New York Times, March 7, 1987)

(Burros goes on to say that an overripe durian smelled terrible.) Perhaps a good counterexample to putridity is the fact that children, known to abhor putrid foods, adore durian. But whether the objective durian has a strong or sweet odor is almost immaterial, as our point in this analysis is that its value in the West changed over time for no physical reason.

10. The Swahili are a Muslim people who live on the East African coast. In some ways they differ significantly from other East Africans. One important distinction for the present discussion is that the Swahili have always typically lived in towns rather than in dispersed farming settlements (like the Kamba or the Luhya) or in nomadic groups (like the Maasai or Samburu). In December 2001, UNESCO granted Lamu’s Old Town “World Heritage” status citing the fact that the Old Town had “retained traditional functions for over 1,000 years” (UNESCO World Heritage Centre).

11. There are two beverages called kahawa ‘coffee,’ kahawa chungu (kahawa tungu in the Lamu dialect), and kahawa tamu. Kahawa sellers often set up on the side of a street, with the customers sitting on rough low benches, sometimes with tables as well. Kahawa sellers are self-contained units.

12. Also called murungi or gat. Sold from street stands, kiosks, or bundles laid out on the ground. Proponents claim it is a mild stimulant, akin to coffee in effect. Opponents claim that it is addictive and akin to amphetamines. The active ingredient is delivered by chewing the stems of branches of a tree heavily cultivated in Meru District in Kenya. One swallows one’s saliva, but spits out the chewed-up stem. Because miraa is very bitter, a chewer keeps something sweet in his mouth while chewing it. That used to be peremende, ‘hard candy.’ Now Big G chewing gum is the favorite. One also sips coffee, or, commonly, punches a nailhole in the bottlecap of a soft
drink and takes little swigs through the hole. There are two main kinds of *miraa: kangeta* and *giza*. *Miraa* is chewed at home as the focal substance of a social conversation event. It is said to stimulate conversation and clarity of thought, as well as to help keep one awake. (Long distance truck and bus drivers chew *miraa* to stay awake and alert.) Many Swahili social events entail staying up all night (for which Swahili has a verb, *kesha*). Street use of *miraa* is not considered respectable. It is chewed sitting on the street by the *miraa* sellers and chewers who socialize with them. Hangout groups of young men may also chew *miraa*. *Miraa* has a bad aura in the popular mind and is associated by some with smokers of marijuana and shiftless persons. During the fieldwork, Kenya *miraa* was legal. A few years ago it was declared illegal in Somalia, where it is extremely popular. Exported to the Middle East from Kenya and Ethiopia, it is also grown in, and quite important culturally in, Yemen.

13. The sociolinguistics of the situation are that politeness dictates that passersby nearing the groups must ritualistically verbally greet the people in them. My own field experience has shown many times that if these greeting rituals demonstrate that the passing strangers speak Swahili, they are often sincerely welcomed – even if they are Euro-American foreigners – to break the fast with the group. (“Sincerely,” meaning that the group members use not just the word *karibu* ‘welcome’ with the common pro forma intonation, but with appropriate intonation and additional phrases.) Even if the strangers demur, saying that they (are undeserving because they) had not fasted that day, they are still welcomed.

14. In an upcoming revision of the present hypothesis, I will hypothesize different meanings. I have shown in the present paper how, in the normal course of events, privacy or, let us call it, insulation from intimacy has gone hand in hand with the garnering of *heshima*, and I have analyzed them as one and the same. But we learn that in this Ramadhan context, *heshima* is accomplished not by insulation from intimacy but from intimacy. *Heshima* is the goal. It can in many contexts be accomplished through privacy, but it is not the same as privacy. Therefore, in the upcoming version of this analysis, I will hypothesize that the privacy – intimacy scale behaviors are meanings of opposition and *heshima* a message that is inferrable from those meanings.

15. But see note 14.

**References**


Index of names

A
Alarcos Llorach, E. 132, 138, 140
Allen, James de Vere 317, 332
Allen, James F. 148–149, 159–160
Anderson, A. 160
Anderson, E. 306
Anderson, J. 223
Andrews, E. 71, 83, 222
Aoki, H. 242, 261
Armstrong, D. 285, 294, 306
Ashton, E. 212, 222, 224–225, 230–237
Avanesov, R. 108, 129
Azim, A. 65, 83, 107, 129
B
Barlow, M. 178, 194
Baron, N. 286, 306
Beckman, M. 95, 104
Belić, A. 208
Bhaskararao, P. 296, 307
Biryla, M. 108, 129
Bloomfield, L. 5, 289, 307
Boomer, D. 98, 104
Boranić, D. 207–208
Bourdieu, P. 317, 332
Brinton, L. 266, 280
Browman, C. 171, 174, 299, 307
Buk, E. 82–83
Bwenge, C. 219
Bybee, J. 170, 174
C
Cekman, V. 108, 129–130
Chafe, W. 255, 261
Chao, Y. 165–166, 174
Cherry, C. 289, 307
Chomsky, N. 5, 26–27, 36–37, 39, 50, 56, 61
Clark, C. 83
Clark, H. 144–145, 160
Cohen, D. 178, 194
Cohen, E. 83–84
Cole, D. 222
Collier, P. 280
Collier, N. 12, 143
de Jonge, B. 10, 12, 14, 65, 131, 140–141, 279, 281
de Saussure, F. See Saussure, F. de
Deacon, T. 301, 307
Dekker, A. 10, 65, 131, 140
Denes, P. 294, 307
Dew, D. 298, 305, 307
Dodg, B. 294, 307–308
Dolgin, J. 316–317, 333
Donley-Reid, L. 317, 333
Dordević, R. 208
E
Eccardt, T. 11, 283
Eco, U. 286–289, 302, 307
Edington, D. 174–175
Edmondson, J. 178, 194
Ellis, J. 310, 333
Enbe, C. 69, 84
Ezrati, R. 84
F
Farsi, S. 318–319, 333
Fathi, A. 65, 84, 107, 129
Fischer, A. 306–307
Fischer, K. 144, 160
Fischer, O. 307
Flores, N. 65, 84, 107, 129
Fontanillo, E. 296, 307
Fowler, C. 299, 301, 305, 307
Fradkin, R. 109, 129
G
Gabrioud, B. 301, 307
Gamkrelidze, Th. 116, 129
Gan, R. 69, 84
García, E. 9, 13, 195, 208
Ghaiduan, U. 317, 319–320, 333
Gibson, J. 305, 307
Gildin, B. 41, 62
Gilliéron, J. 20, 39
Givón, T. 240–241, 261
Goody, J. 333
Gorup, R. 8, 41, 62, 192, 195
Grabtshikov, S. 108, 130
Green, H. 69, 84
Greenbaum, S. 160, 281
Greenberg, J. 129, 150, 160, 214, 222
Grice, H. 264, 280
Grunwell, P. 69, 84
Guthrie, M. 214, 222
H
Haas, W. 211, 222
Halliday, M. 252, 261, 265–266, 278–280
Hameed, S. 65, 84, 132, 139, 140
Salmon, J. 83, 85
Sanders, G. 221, 223
Sapir, E. 71, 74, 85
Schein, J. 305, 308
Schiffrin, D. 144, 161, 264, 266, 281
Schneider, D. 333
Schwenter, S. 266, 281
Sebeok, T. 287, 289, 308
Shaumyan, S. 305, 308
Shiroti, S. 242, 262
Shokty, L. 81, 83, 85
Sokolov, R. 310, 334
Stampe, D. 69, 85
Stemberger, J. 90, 96, 105
Stevanović, M. 209
Stewart, D. 305, 308
Stoel-Gammon, C. 96, 105
Strobel, M. 318, 334
Sudnik, M. 108, 130
Sudnik, T. 108, 130
Summerfield, Q. 293, 308
Sussex, R. 108, 130
Svartvik, J. 160, 281
Sweetser, E. 265–266, 281
Teramura, H. 242, 262
Tsutsui, M. 242, 249, 258, 261
Tubul, R. 83, 86
Twaddell, W. 166, 175
V
Van Hoek, K. 181, 194
Vitale, A. 224, 235–237
Vyhonnaia, L. 108, 130
W
Wald, B. 329
Wallace, A. 312, 331, 334
Ward, N. 145, 161
Waugh, L. 52, 62, 71, 85–86
Welmers, W. 218, 223
Wexler, P. 108, 110, 130
Whitehead, H. 325–326, 334
Willis, H. 264, 281
Wolkomir, R. 311, 334
Wundt, W. 285, 308
Z
Zipf, G. 51
Zribi-Hertz, A. 177, 181, 189, 194
Subject index

A

a priori
dative as 235–236
in anthropological theory 325
in linguistic theory 34, 39, 43–57, 60–61, 236, 239
reflexive as 177, 181–182, 192, 207

accusative case, see case; Control, System of Degree of

acoustic frequency, see frequency, acoustic

acquisition, see language acquisition, see also learner

Additional Articulator Hypothesis, see articulator, additional

adjectival prefix 212–213, 216, 218

adjective 22, 52, 218, see also adjectival prefix; stem, adjectival

adjective declension 74–75, 78

adjective-noun order 22, 52

agreement 216–218

adroitness of articulator, see articulator, adroitness of

agent 200–206, 224, 237

agreement 215–218, 221

abu 314–315, 323, 325

allophone 64–65, 165, 173–174

alphabet 65, 167–169

alternation
formal alternation 213, 215–219

sound alternation 99

American English, see English, American

American Sign Language 283–284, 286, 294, 297, 299, 301, 323

anterodorsum 65–66, 110, 113–114, 122, see also articulator

anthropology 325, 330

anti-apriorism, see apriorism

anti-nomenclaturism 5, 31, 32, 35

aperture 65, 68, 94, 96, 46, 111–112, 114, 117, 124, 139, 171


aphasia 89–90

apical, see apex of the tongue

apricotism 45–46, see also a priori

Arabic 83, 220

arbitrariness 283, 289, 292, 310–313, 323, 326, 330

architectural space 314–317, 328


active 65–67, 95, 112, 167

additional, or number of 65, 67–68, 72–76, 79, 82, 89, 95, 119–124, 132–133, 172

adroitness of 66–67, 72, 76, 79, 95, 118, 122, 138, 139, 146, 172

reuse of 68, 110, 124–127, 172

articulatory gesture, see gesture, articulatory

articulatory subgesture, see subgesture

ASL, see American Sign Language

assimilation 88, 95–96, 306

Attention, System of, see Deixis, System of

augment 214

augmentative 216, 218–221

B

Bantu 212, 214–215, 217, 222, 236, 237

Proto 214–215, 222

Proto Northeast Coast (PNEC) 214, 222


binary feature 53

binding theory, see Government-Binding Theory

bootstrapping approach 44–46, see also minimalism in Columbia School, zero responsibility

Byelorussian 107–130

C

case 26, 51–52, 56, 74–79, 195, 200, see also Control, System of Degree of; dative case

Focus


phonological 65, 128, 150

message categories or categories of use 183, 207, 224, 231–232, 234–236

cavity 111–112, 114–115, 119, 133, 139, 295, 298

channel 293, 295, 300, 302

child phonology, see phonology, child

chi-square test of significance 10

class, see noun class, social class

clinical phonology, see phonology, clinical

clicit 195, 200, 206–207, 235

co-articulation 68, 72, 299–300

Cognitive Grammar 2, 5–6, 36, 52, 163, 181, 260

communication 2, 7, 35, 57, 93, 107, 118, 124, 144, 147, 149, 159, 168, 169, 198, 247, 284, 287–289, 292, 296, 327–329; see also communicative behavior, communicative distinction, communicative force, communicative goal,
communicative orientation, communicative strategy intended 186, 188
maximum, with minimal effort 64–66, 68, 74, 77, 80, 88, 100, 107, 121, 128, 146, 173
communication factor, see communicative orientation
communication load, see communicative force
communicative behavior 287, 289–291, 297, 305
communicative distinction 66–67, 72, 76, 89–91, 96
communicative factor, see communicative orientation
communicative force 67–68, 72, 82, 89, 91–93, 100–103, 118, 120–121, 134, 150, 172–173
communicative goal 266–269, 276–277
communicative load, see communicative force
communicative opposition, see communicative distinction
communicative purpose, see communicative goal
communicative strategy 3, 7, 10
Complexity of Articulation 132–134, 136–139, see also articulator, additional compositional view of meaning, see meaning, compositional com-sign 290, 302
conjoined expression 179, 181–182, 190–191, 193
conjoined information 264, 266, 268, 276, 278
conjunct 190–191, 264, 266–267, 269–270, 272, 275–279
conjunction 248, 249, 253–257, 269–270, 272, 274, 278
cluster 68, 72–73, 88, 90, 96, 100, 114–117, 119–121, 124–126, 132, 139, 151
construction 9, 45, 187, 224, 230–231
nonlinguistic 314, 324, 327, 329, 332
phonetic 97
continuity, see discreteness versus continuity in semiotics
Control, System of Degree of 21, 197, 200–207, 226–235
coronal 95
croatian 195, 207, see also Serbo-Croatian
Czech 211
D
dative case 26, 224, 235, see also case; Control, System of Degree of decoder 88–89, 118
deductive 51, 56–61
deduction 88–92, 101–103, 306
derivation 212, 220–221
descriptive linguistics 10, 163–164, 173
diacronic 63, 65, 69, 82, 110, 116
differentiation, System of 22
diminutive 216, 218
discourse 3, 5, 10, 35, 54, 143–150, 154–155, 157, 163, 177, 181, 188–189, 214, 266
Discourse Referent, System of 3, 197–200
discreteness of grammatical or sign systems 25, 33, 35
versus continuity in semiotics 283, 292–293, 295, 305
distinctive feature 90, 91, 96, 100, 114, 115, 325
distortion 298–301
double articulation 297
durian 312–313, 330–331
Dutch 19, 91
E
echo location 295
empty morph 196–197
encoder 88–89, 118
American 145, 149–151, 273
Glaswegian 149–151
Middle 78–80
Modern 71–74, 77–80
Old 63, 71, 74–81
Scottish 151, 157
epistemic 240, 259, 265
Event Focus, System of, see Focus
evidential 239–241, 243, 245, 355
evolution of language, see historical change
explosion (of stop) 115–116, 133, 139
F
faithfulness, constraints of 89
feedback, speech 293–295, 303
figurae 82, 285
Focus
Event Focus, System of 41
Focus on Type of Inference, System of 247, 250, 258
Participant Focus, System of 41, 197–207
food 309–315, 318–326, 328, 330–331
formant 121, 298
Form–Content School 41
French 19–20, 41, 52, 235, 306
frequency, acoustic 48, 295, 297
gender, see also Sex, System of grammatical 71, 74, 78–79, 195, 197–200, 202
social 314, 316, 329
generative grammar 2, 37, 42, 53–54, 56–57, 60, 163, 179, 181, 188, 224, 235–236
generative theories of phonology 69
gestural overlap 299–300
gestural score 171
gesture, see also gestural overlap, gestural score, subgesture, trace of gesture articulatory 74, 88, 110, 115–117, 120, 124–125, 128, 132–133, 138, 146, 170–171, 283–306
semiotic 284–287, 293–298, 301–306
Glaswegian English, see English, Glaswegian
Glottis 66, 113, 119, 121, 133, 171–172
Government-Binding Theory 56, 179, 181
grammar 4, 13, 39, 77, 207, 215, 224, 235, 244, 279
Cognitive, see Cognitive Grammar
Columbia School 3, 9, 50, 53, 59, 164, 279
discourse 188
generative, see generative grammar
Jakobsonian 53
Universal, see Universal Grammar
grammatical system 3, 10, 12, 21–26, 32–35, 37–38, 73, 207, 263, 279, 309
grammatical tradition 26, 36–37, 45, 51
Greek 3, 9, 21, 26, 51–52, 169–170
Guillaumean School 2, 42, 53–54

honofric 257–258
Hungarian 83, 145

I
impersonal 195–197, 207, 221
index 291
inductive 42, 44, 51, 53, 57–59
Inference, Type of, System of 246–251, 253–259
inflectional morphology, see morphology, inflectional information loss 291, 298–299
insistence on a referent 177–178, 187–188, 190–191
instrumental meaning, see meaning, instrumental interchangeability 302
intonation 69, 152, 302, 305, 332
irrealis 240
irregular inflectional morphology, see morphology, irregular inflectional
Italian 19–20, 41, 107, 116, 118, 126, 139, 166, 279
J
Jakobsonian school 2, 42, 52–54, 325
K
Kind of Differentiation, System of 22–23
kinesthetic 293–294, 297, 302–303
L
Ladino 83
language acquisition 81, 88, 91, 294, 303, 306
language 3–4, 17–21, 23–25, 27–39, 45, 49, 60, 64
larynx 113–114, 116, 119, 133, 296
Latin 19, 21, 26, 51–52, 83, 116, 118, 182, 235
law, scientific 58, 60
learner 49, 215, 218, 303, 323, 326
lexical class 218, see also gender, grammatical lexicon 3, 9–10, 12, 39, 50, 69, 71, 101, 131, 169–171, 278–279
linguistic sign theory 42, 60, 329
lip 65–66, 76, 95, 114, 121, 140, 146, 151, 164, 294, 296, 298, 301, 304
lip-reading 293–294, 302, 305
loanword 87–88, 99–103, 219–221
logical operator 264
logic-based structure 51, 57
M
maleficary 225, 232–233, 235
Mandarin 166
markedness 3, 13, 53, 54, 89, 116, 150–151, 153–157, 163
marked or unmarked sound or phone 67, 71, 75–76, 80–82, 90, 94–95, 138, 146, 150–151, 154–155, 157–158
marked or unmarked
form 75, 195, 213
McGurk effect 293–294
compositional 3, 38, 45, 54–55, 57
instrumental 22, 34, 54–55
lexical 3, 69–70, 205, 224–225, 263–267, 278
meaning-message distinction 3, 8–10, 22–23, 27, 34, 38, 197, 223, 235, 245, 324, 328
nonlinguistic 13, 302, 309–330, 332
sentence 3, 23, 35, 38, 45
Middle English, see English, Middle
minimalism in Columbia
School 46, see also bootstrapping approach, zero responsibility
mobile 66, 68, 82, 110, 114–117, 120, 122, 124
Modern English, see English, Modern
modification grammatical 52, 221
sound 295–296
modifier 221
monosemy 264
mora 97–98, 103
morpheme 22, 47, 72, 81, 164, 172, 214, 220, 285, 289, 299–302, 304
inflectional 71, 74–75, 80
zero 21–22, 48, 72–73, 75, 78, 211, 213, 215, 220–222, 328
morphology 3, 21, 37, 170, 228, 230, see also morpheme
derivational 81, 83
inflectional 63, 71–82
irregular inflectional 73, 77, 79
morphophonemics 214, 216–217
multisensory perception 294
N
Niger-Congo 218
noise 47–48, 289, 293, 295–296, 298, 304
nomenclaturism, see anti-nomenclaturism
noun class 211–221
number grammatical 71, 74, 78–79, 82, 195, 212–215, see also Number, System of of articulators, see articulator, additional of phonemes 65, 67, 82, 91, 93 of syllables 92, 151, 155
Number, System of 9, 21, 38, 187, 197–200, 202, 309
O
observation 13, 35–36, 38, 46, 50, 57, 59–60, 168, 173, 327–328
Old English, see English, Old
opposition binary 45, 47
communicative 72–73, 75–76, 89–91, 96
lexical 279
morphological 221
of exclusion 3, 54
of inclusion 54
of substance 13, 195, 197
of value 3
phonological or phonemic 46, 66, 150
semantic 246, 264, 329, 332
spanned 56
Optimality Theory 69, 88–89
orientation 10, 59–61, 64, 173, 302, 327, see also communicative orientation, human factor, phonetics
ostentation 286
P
palatalization 108, 110, 122
parole 3, 18, 49, 64
part of speech 3, 36, 51–52, 73–74, 83, 143, 149
participant 8, 147–149, 195, 197, 199–207, 226, 228, 230, 233–234, 257
Participant Focus, System of, see Focus 195, 197, 199, 207
passive receptor 65–66, 72, 79, 110, 112, 122, 128, 167
passive voice 196, 207
periodic table of the elements 58
person, grammatical 187, 195, see also Discourse Referent, System of
pharynx 66
PHB, see Phonology as Human Behavior
and allophone 64–65, 173–174
moraic 97–99
number of 65, 67, 82, 91, 93
supra-segmental 48
phonemic principle 164, 170
phonological unit 164, 167, 172, 288–289, 302
acoustic 304
articulatory 171, 302
child 94–95, 98–99, see also developmental
clinical 63, 69, 89–91, 107
developmental 63, 69, 89, 91, 93, 107
diachronic 63
direct realist theory 302
Natural Phonology 69
Optimality Theory 69, 88–89
Prague School 63
phonotactics 67–69, 81, 107, 128, 145, 220
place of articulation, see passive receptor
Plains Indian Sign Language 286
PNEC, see Bantu, Proto Northeast Coast
polysemy 45, 55, 265, 266
posterodorsum 65–66, 113–114
postvocalic r 316
pragmatics 35, 178, 245, 251, 264–265
Prague School 45, 63, 150
prefix 211–222, 258
preposition 52, 77, 79, 224, 230, 235
presentation (method of referring) 286
prestige 314, 316
Probability, System of 21, 41, 237
pronominal, see pronoun
pronoun 197–198, 200, 212, 214–215, 221, 252
prosody 48, 69, 99, 151–152, 157
Proto Bantu, see Bantu, Proto
Proto-Germanic 80–81
Proto-Indo-European 80–81
prototype 181
R
radio 291, 293
Ramadhan 319, 323–324
reduplication 92, 224, 231
redeployability in communicative behavior 302
respect 314, 328
reuse of articulator, see articulator, reuse of
Romance 83, 116, 195
rule 3, 26, 37, 144, 217–220, 315
cultural or societal 310, 314, 318, 321, 324
Russian 82, 110, 116, 121
S
scene 55, 183, 202–203, 224
science 18, 25–26, 34, 35, 42, 44, 54, 57–60
Scottish English, see English, Scottish
secondborn 303, 306
segmentation 285, 292, 299
semantic domain, see semantic substance
semantic substance or domain 10, 12, 21–22, 25, 32–33, 164, 195, 197, 226, 246, 263, 266, 279, 309, 314, 327
semantic universal 54
semi-consonant 295
semantic 18, 63–64, 81, 88, 163, 283–292, 303, 305
semi-vowel 65, 112, 121, 295
sentence grammar, see grammar, traditional
Serbian 207, see also Serbo-Croatian
Serbo-Croatian 41, 195–207, see also Croatian, Serbian
Sex, System of 187
shame 314
sign ASL 297, 323, see also American Sign Language, sign language
linguistic or speech 2–4, 9, 20, 24–25, 27–38, 49, 60, 80–82, 163, 197, 283–285, 287, 288, 292, 294, 296–297, 299, 302–305, 323, 329, see also signal-meaning pair, signe linguistique
semiotic 283–289, 291–293, 295–297, 301–303, 323
sign-based or -oriented 2–4, 17, 34, 38, 42, 49, 51–52, 54, 60, 71, 107, 245–246, 263, 278
sign system 23–25, 33, 35–36, 63–64, 329
sign theory, linguistic 42, 60, 329
sign language 283, 284–286, 293–299, 301, 303, 323
signal acoustic or sound 283, 296–297, 299, 302
light 296–297
linguistic 3, 6, 9–10, 12, 21, 22, 26, 28, 30, 32–34, 47–48, 51, 60, 81–82, 144, 163–164, 166–168, 170, 172–173, 187, 195, 197, 200, 223, 228, 263–264, 267, 288–298, 326–327, 330, see also signal-meaning pair, signifier
paralinguistic 144
speech 64
signal-meaning pair 17, 45, 246, 323, 329, see also sign, linguistic; signe linguistique
signe linguistique 4, 17, 28, 32, 34–35, 38, 163, see also sign, linguistic
signifiant, see signifier, see also signal
signification 30, 32, 35, 38, see also meaning, signified
signifié, see signified, see also meaning, signification
signified 32, 34, 38–39, 49, 81, 284–285, 291, 302, 326, see also meaning, signification
signifier 32, 38–39, 49, 80–81, 283–285, 287–289, 291–294, 326, see also signal, linguistic
Sino-Japanese 95
social class 315–316, 318–323
sociolinguistics 107, 144, 310, 316, 332
sound image 288, 299
sound wave 5, 38, 46–48, 53, 57, 286, 289, 291–292, 294, 327, 328
space 309, 314–321, 328
Spanish 41, 83, 107, 131–140, 296
specialization 287, 291
spectrogram 292, 295, 298–301, 304
speech chain 294, 296–297, 300
speech error 87, 89, 92–96, 98, 303
stable 66, 68, 79, 82, 110, 114–117, 122, 124–125, 128
stem
adjectival 216–218, 220
lexical 72–73, 80, 217, 224, 231	noun 212, 218
verbal 220, 233, 236
stimulus 289, 291, 300, 302
stone house 317, 320–321
strategy, see communicative strategy
stricture 65, 76, 82, 96, 111–114, 117
stroke 284–285, 302
structuring cue 267, 278–279
substance 3–4, 11, 18, 20, 24–25, 27–33, 35, 38, 164, 169, 211, 305, 310, 313, 330, see also opposition of substance, semantic substance
substitution 75, 88–91, 93–95, 99
suffix 72–73, 80, 101, 225, 231, 260
suppletive 217, 219
supra-segmental 48, 166
Swahili language 211–222, 223–237
society 309, 313–315, 317–320, 323–324, 331–332
syllable position hypothesis 98
symbol 283, 285, 289–296, 300, 302–305
symptom 287, 291, 302
synchronic 65, 69, 116, 214, 220
T taboo 319, 325–326
textual structure 264, 278
thematic relevance 266–269, 272, 276–279
Theory of the Sentence 52, 57, see also grammar, traditional
Time, System of 3, 21, 32
trace of gesture 284, 286, 291–292
traditional grammar, see grammar, traditional phonetic and phonemic categories 65, 112, 114–115, 128
semiotic categories 287, 291–292
transition in articulation 68, 82, 93, 111, 124, 295
truth condition 183
turbulence 68, 111–112, 117, 304
Type of Inference, System of 247–248, 250
U universal 7, 10, 47, 52–54, 150–151, 327
Universal Grammar 36, 56
unmarked, see markedness
uvula 66, 111, 121
V validation 3, 6, 10, 54
Validity of Claim of Position, System of 314
verbal particle 196–197
postverbal particle 223
visible gesture or articulation 286, 294, 297
vision 118, 293, 295
vocal cord or fold 66, 72–73, 76, 79, 170, 171, 173, 295, 296, 298–300, 306
voice grammatical 196, 205
Z zero
morpheme, see morpheme, zero
responsibility 7, 48, see also bootstrapping approach, minimalism in Columbia School
In the series *Studies In Functional And Structural Linguistics* the following volumes have been published thus far or are scheduled for publication:

44 JESSEN, Michael: Phonetics and Phonology of Tense and Lax Obstruents in German. 1999. xx, 394 pp.