PROCEEDINGS
OF THE CONFERENCE
ON COMPUTER-RELATED SEMANTIC ANALYSIS

Held under the auspices of
WAYNE STATE UNIVERSITY
at Las Vegas, Nevada
December 3-5, 1965

D D C

JUL 2 5 1967

C

PAGES _____ ARE MISSING IN ORIGINAL DOCUMENT

FOREWORD

The first of the "Princeton-type" meetings in the field
of machine translation took place July 18-22, 1960, and was
devoted primarily to the task of bringing together various
groups working in machine translation in order to exchange
information of mutual interest to them. The second conference
in this series took place April 4-7, 1961, at Georgetown
University in Washington D.C., and was devoted to problems
of grammar coding. The third meeting was held June 13-15, 1962,
at Princton, and dealt with syntax. The fourth meeting in this
series was held at Las Vegas, Nevada, December 3-5, 1965,
immediately following the Fall Joint Computer Conference and
dealt with problems pertaining to Computer-Related Semantic
Analysis.

The agenda committee for this conference consisted of
Harry H. Josselson, Wayne State University, Martin Kay, RAND
Corporation, Susumo Kuno, Harvard University, and Eugene D.
Pendergraft, University of Texas. The Conference was housed
at the Sands Hotel, Las Vegas, Nevada. Registration started
Friday afternoon, December 3. There was an evening meeting,
at which a keynote address was delivered by Professor Winfred
Lehmann of the University of Texas, President of the Association
for Machine Translation and Computational Linguistics. There
were two sessions on Saturday, December 4, and one session
Sunday morning, December 5, with the Conference terminating
at noon of that day. Scholars with experience in semantic
analysis and/or computer processing of semantic data were
invited to address the meeting.

In addition to the keynote address, thirteen papers were
presented at the meeting. Besides discussions immediately
following the presentation of papers, two sessions for informal
discussion were held on Friday and Saturday evenings. Of the

six foreign scholars addressing the conference, three were
from the United Kingdom and one each from Hungary, Italy,
and Israel. In addition to observers, eleven federally
sponsored groups engaged in research in automatic transla-
tion and related areas were represented at the conference.
Also in attendance were representatives of several interested
U.S. Government agencies.

As indicated above, prior to the Las Vegas meeting on
"Computer-Related Semantic Analysis," MT research groups
supported by U.S. federal funds met on three previous
occasions in order to discuss problems of mutual interest.
The proceedings of these meetings (at Princeton, New Jersey
in 1960 and 1962, and at Georgetown University in 1961) were
published in mimeograph form and were distributed by Wayne
State University, under whose auspices these conferences were
held, among the conference participants and all others
interested. The main objective of these reports was the
dissemination of information to interested groups and indi-
viduals who were unable to participate in or attend the meet-
ings, since for working purposes, attendance at the conference
had been restricted to representatives of federally sponsored
MT groups.

In keeping with this tradition, a mimeograph report of
the proceedings of the Las Vegas meeting is being published.
This report includes all papers presented at the conference
as well as a summary of the discussions which followed each
presentation. The latter have been edited primarily for style
and elimination of repetitious material. The proceedings are
available to all interested.

Thanks are due to the National Science Foundation, the
Office of Naval Research and the U.S. Air Force for financial
support of this conference. Appreciation is also expressed
herewith to Wayne State University for convoking this meeting,

to the Agenda Committee for drawing up a thought-provoking
program, and to the participants who contributed to the
success of the meeting by their presentations and discussions.

Harry H. Josselson
Department of Slavic and
  Eastern Languages
Wayne State University
Detroit, Michigan 48202

TABLE OF CONTENTS

Participants at the Conference on

Computer-Related Semantic Analysis

held at the Sands Hotel in

Las Vegas, Nevada

December 3-5, 1965

SPEAKERS:

Yehoshua Bar-Hillel
Hebrew University, Jerusalem

Elinor Charney
Massachusetts Institute of Technology

Ernst von Glasersfeld
Instituto de Documentazione,
Associazione Meccanica Italiana

Ferenc Kiefer
Computing Centre of the Hungarian Academy of Sciences

Winfred Lehmann
University of Texas

Margaret Masterman
Cambridge Language Research Unit

Eugene D. Pendergraft
University of Texas

Warren Plath
IBM, Yorktown Heights

Herbert Rubenstein
Center for Cognitive Studies, Harvard University

A. Kimball Romney
Harvard University

John Ross
Harvard University

Karen Sparck Jones
Cambridge Language Research Unit

SPEAKERS: (cont.)

Stephen Ullmann
University of Leeds

Uriel Weinreich
Center for Advanced Study in Behavioral Sciences

REPRESENTATIVES OF FEDERALLY SPONSORED GROUPS:

Bunker Ramo Corporation   - Gerhard Reitz, R. M. Worthy

Georgetown University   - Ross MacDonald

Harvard University   - Susumo Kuno, John Ross

IBM, Yorktown Heights   - Dorita Lochak, Warren Plath

Informatics Inc.   - Jules Mersel

RAND Corporation   - Kenneth Harper, David G. Hays

University of California
    at Berkeley   - Douglas Johnson

University of Chicago   - Victor Yngve

University of Indiana   - Robert Wall

University of Texas   - Eugene D. Pendergraft, Wayne Tosh

Wayne State University   - Leon Bruer, Sidney Simon

SPONSORING U.S. GOVERNMENT AGENCIES:

National Science Foundation   - Eugene Pronko

Office of Naval Research   - Gordon Goldstein

US Air Force, Griffiss A.F.B.   - Zbigniew L. Pankowicz

REPRESENTATIVES OF U.S. GOVERNMENT AGENCIES

U.S. Air Force, ESD   - Lt. Bruce Frazer

U.S. Air Force,
Office of Scientific Research - Rowena Swanson

U.S. Government   - Herbert Avram

PRESIDENTS OF THE ASSOCIATION FOR MACHINE TRANSLATION AND COMPUTATIONAL LINGUISTICS:

Paul Garvin   - president elect

David G. Hays   - past president

Winfred Lehmann - president

ARRANGEMENTS COMMITTEE:

Harry H. Josselson, Chairman

Martin Kay

Susumo Kuno

Eugene D. Pendergraft

OBSERVERS:

Russel Abbott   - General Electric, Tempo, Santa Barbara

Colin Bell   - IBM, Yorktown Heights

Susan C. Berezner   - General Electric, Tempo, Santa Barbara

Daniel G. Bobrow   - Bolt, Beranek, and Newman Inc.

Hannah K. Cole   - University of Texas

John C. Fisher   - General Electric, Tempo, Santa Barbara

Peter Ingerman   - Radio Corporation of America

OBSERVERS: (cont.)

Robert Langevin	- Technical Operation Inc.
John Olney	- Systems Development Corp.
Seymour Pappert	- Massachusetts Institute of Technology
Anne Saporito	- U.S. Government
John Robb	- Consultant, Washington,D.C.
K.G. Schweissthal	- Forschungsgruppe LIMAS, Bonn, Germany
Robert Simmons	- Systems Development Corp.
Percy Tannenbaum	- Center for Advanced Study in the Behavioral Sciences
Donald Walker	- Mitre Corporation

Program of the Conference on
COMPUTER-RELATED SEMANTIC ANALYSIS
held at the Sands Hotel in
Las Vegas, Nevada
December 3-5, 1966

FRIDAY, DECEMBER 3

6:00- 7:30 p.m.	Informal discussion	Eugene D. Pendergraft University of Texas Chairman
9:30 p.m.	Keynote Address Winfred Lehmann University of Texas	INTERFACES OF LANGUAGE

SATURDAY, DECEMBER 4

9:00-10:30 a.m.	SESSION I	David G. Hays RAND Corporation Chairman
	Yehoshua Bar-Hillel Hebrew University Jerusalem, Israel	THE OUTLOOK FOR COMPUTATIONAL SEMANTICS
	Uriel Weinreich Center for Advanced Study in Behavioral Sciences	SOME TASKS FOR SEMANTICS
10:45 a.m.- 12:45 p.m.	Elinor Charney Massachusetts Institute of Technology	STRUCTURAL SEMANTICS: THEORY OF SENTENTIAL MEANING
	Margaret Masterman Cambridge Language Research Unit England	SEMANTIC ALGORITHMS
2:30 p.m.	SESSION II	Winfred Lehmann University of Texas Chairman
	Stephen Ullmann University of Leeds England	SOME QUANTITATIVE PROBLEMS IN SEMANTICS AND LEXICOLOGY

SATURDAY, DECEMBER 4

SESSION II

Herbert Rubenstein   PROBLEMS IN AUTOMATIC
Center for Cognitive   WORD DISAMBIGUATION
 Studies
Harvard University

4:15- 5:45 p.m.   Ferenc Kiefer   SOME SEMANTIC
Computing Centre of the   RELATIONS IN NATURAL
Hungarian Academy of   LANGUAGE
Sciences
Hungary

John Ross   UNDERLYING STRUCTURES
Harvard University   IN DISCOURSE

A. Kimball Romney   MULTIDIMENSIONAL
Stanford University   SCALING AND SEMANTIC
                          DOMAIN

6:00- 7:30 p.m.   Informal discussion   Martin Kay
                                   RAND Corporation
                                   Chairman

SUNDAY, DECEMBER 5

9:00 a.m.   Session III   Paul Garvin
                                     Bunker-Ramo
                                      Corporation
                                     Chairman

Karen Sparck Jones   SEMANTIC CLASSES AND
Cambridge Language   SEMANTIC MESSAGE FORMS
Research Unit
England

Eugene D. Pendergraft SEMANTIC SELF-
University of Texas   ORGANIZATION

10:45 a.m.-   Warren Plath   A TAG LANGUAGE FOR
    12:15 p.m.   SYNTACTIC AND SEMANTIC
                                     ANALYSIS

Ernst von Glasersfeld AN APPROACH TO THE
Instituto de   SEMANTICS OF
Documentazione,   PREPOSITIONS
Associazione Meccanica
Italiana
Italy

KEYNOTE ADDRESS

INTERFACES OF LANGUAGES

Winfred Lehmann
University of Texas

Among reasons for the meeting of linguists with specialists
in the computer sciences is the assumption that each will contri-
bute to the interests of the other. The contributions must be
more than superficial. Linguists hope to obtain increasingly
complex dissection of texts from computers, but they might do
so with no understanding of computational theory. Certainly
there will be increasing use of computers by scientists,
humanists, and apparently even housewives, who will understand
them as little as they do an automobile or a typewriter. But
presumably there may be closer parallels between the function-
ing of a computer and the behavior of a speaker than there are
between the operation of a typewriter and the cerebral activity
of the scholar or writer using it. More accurately, a closer
relationship may be achieved between the programs developed
for computers and man's language. Earlier this year I discussed
some of the comforts, possibly even results, we can derive by
examining this relationship; see "Toward Experimentation with
Language," Foundations of Language 1 (1965) 237-248. I have
now been jostled into a further statement, and, as you know from
my title,would like to discuss similarities in the mechanics
of interrelating separate computer programs with those used
by linguists to manage components which they may handle as
distinct.

Beyond their clerical contributions, the great interest
which computers have for us is their capability of simulating
language. We can provide them with linguistic data and let
them generate phonological entities: Bengt Sigurd discusses

results of such activities for Swedish in his recent _Phonolo-gical Structures in Swedish_ (Lund, Uniskol, 1965). We can also have computers produce syntactic entities, as a number of you have done. In _The Graduate Journal VII_ (1965) 111-131, I have reported on the Linguistics Research Center translation experiment of January 1965. Clearly these products of the computers have not simulated language, but only fragments of it. In proceeding to the simulation of language we must not only aim at fragments, but at language as a whole. But how can we arrive at such an achievement.

In considering the achievements of computational linguis-tics to the present we have dealt primarily with our inadequate understanding of language. Virtually any statement on semantics or meaning tells us that linguists cannot handle it, or even that they scarcely know where to begin the activities which will lead to its management. Another recent restrained comment has been published in a book that is not otherwise notable for insecurity, Noam Chomsky's _Aspects of the Theory of Syntax_ (Cambridge, MIT Press, 1965). Yet if in our inability to handle meaning we merely string together phonological items or syntactic items with no reference to their meaning, we simulate only a micro-language. To be sure, there have been efforts to move beyond such a narrow segment of language; this meeting is one of them. And some linguists have actually made progress toward a theory which would embrace the meaning component of language. The next few days will tell us more about this progress. But I leave the inadequacies of the linguists for a moment and turn to the programmers and the structure of our instruments of simulation, the computer.

There probably has never been such a well-heeled mysti-fication of mankind as during the last half-generation during which computers have begun their drive to dominate man. Even the auto industry, with its long start in public relations, is unable to match the glittering language under which new

models of computers emerge. On a far lower level, but
relatively a level as exhilarating for academicians as
was the harnessing of the electron for the merchants of
mechanical number bashing, was the possibility of produc-
ing programming languages rather than new theorems for
a journal which would be stored in musty libraries. Only
a Wolcott Gibbs could do credit to the molders of the For-
trans, Cobols, IPL/1's and the like, which are produced,
revised, sub- and super-scripted in such profusion that the
mind of man and machine reels. But when we examine the under-
lying principles of the available computers and their languages,
it is not unjust to say that they are little more than souped
up adding machines which amaze us primarily because they add
with virtually the speed of light. Since adding is a useful
activity in quantitative activities--whether in baking a cake,
changing the course of a rocket, or performing a scientific
calculation--any device which can add rapidly puts us in its
debt. But if we limit our process of communication to addition,
we subject ourselves to various limitations. First, we have a
non-hierarchical system of communication. Second, and less
consequential, we limit the elements of that system--in the
generally used numerical system to ten, in the binary system
of the computer to two. Third, and probably least vital, there
is some interest in having communication systems that are easy
to use.

The first two limitations of communication by computers
need little comment among linguists. No human language has
been restricted to as few as ten elements, nor have any failed
to make use of various hierarchies. Since computer programs,
and computer logic, have been built around a limited number
language, they have been made highly complex in their arrange-
ments, to compensate for the small set of elements and the
lack of hierarchies.

But recently there have been steps toward breaking away
from the simple computer logic and the linearity of program-
ming language. Among these is Licklider's proposed "coherent
programming" -- in which he aims at a "coherent system of
compatible linkable routines," see (Foundations of Language 1,
247-48). In reading about this proposed system a linguist
sees close relationships with natural language. I do not
plan to deal with Licklider's proposal here, but as my topic
indicates I will discuss problems that arise in setting out
to produce a system of linkable routines and of linkable
programs. Links between distinct programs are known as inter-
faces. If computer programs of the immediate future will
come to resemble natural languages the interfaces between
different programs may not be unlike the intermediate layers
between the descriptions of various levels of language.
Programmers may then be interested in learning how linguists
manage these, and linguists may find in the activities of
programmers and computer designers some indications of use-
ful procedures for their own purposes.

In discussing the relationship between description of
language and programming systems we might recall the standard
view of language promulgated in the second quarter of this
century. By this view language consists of various levels,
at least a phonological and a grammatical. The phonological
level in turn consists of a set of signs, which occur in a
language in certain relationships to one another; in English,
for example, t and d are distinct signs because they contrast
with one another, as in sight and side. By a widely used
terminology, t and d are called separate phonemes. Although
these phonemes pattern in this way at the phonological level,
at the grammatical they may pattern differently: when we make
the past tense of rip we use t, though for rib we use d. This
merger at the morphological level may also be exemplified in

verbs like <u>sigh</u>, for which we automatically use <u>d</u>, <u>sighed</u>.
The contrast which exists between <u>t</u> and <u>d</u> at the phonological
level does not exist at the grammatical, and here <u>t</u> and <u>d</u> are
variants of one entity. The difficulty has been handled by
assuming two distinct levels in language: <u>t</u> and <u>d</u> are then
set up as distinct phonemes, but they are variants of one
morpheme. Yet obviously these two levels are components of
one larger system, and accordingly the relationship of their
components to each other must be specified. The specification
has been done by proposing an interface, the so-called morpho-
nemics of our grammars.

Linguists have differed in their views of morphophonemics.
Probably most recent grammars deal with it as a sub-section of
morphemics, or grammar. The linguist who departed most deter-
minedly from this position was Hockett, who in <u>Modern Linguistics</u>,
(New York, Macmillan, 1958) set it up as a separate system of
language, defining it (p. 137) as "the code which ties together
the grammatical and the phonological systems." In much current
linguistic work, on the other hand, the distinction between the
phonological and syntactic component of language is minimized,
and morphophonemics is dealt with partially under the "syntactic"
component, partially under the "phonological." To put this
varying situation in the language of the computer programmer, by
one approach morphophonemics is a partial interface within one
system of programs, by another it is a totally separate interface,
the one in one programming system, the one in the other -- and
the two systems themselves are closely interrelated.

I am not concerned with determining the most economical
linguistic procedure -- with arraying by some kind of value
judgement the various linguistic approaches. I am only interested
in noting how linguists handle one of their problems: the rela-
tionships between various posited levels. If we note this use
of the morphophonemic interface, we find variation in distinct-
ness from approach to approach. Moreover, I haven't forgotten

that this is a conference on semantics rather than phonology
or syntax. But as you have probably assumed, I am following
the notion that the relationship between the phonological and
the syntactic componenets of language is comparable with that
between the syntactic and semantic. If this notion is valid,
the view we accept of interfaces in the phonological area will
have pertinence for our exploration of the semantic, though
it may be simpler to deal with the much explored problem in
this area than with that in semantics and syntax.

The use of interfaces in programming systems may clarify
for us the varying positions. I select illustrations from
systems developed at the Linguistics Research Center under
the direction of Eugene Pendergraft, partly because I know
something about them, but more because you do too, or should
if you have time to undo the packets that neither snow, nor
ice, nor dark of night keep from streaking into your offices.

The Linguistics Research Center's first aim, stimulated
by a penetrating charge from its initial sponsor -- a charge of
ideas as well as dollars -- was study of the possibilities of
translating by computer. In handling this charge a complex
system of programs was produced, now called Linguistic Research
System. By means of LRS various facets of language can be
analyzed and manipulated: currently the grammatical, or in today's
fashionable term, the syntactic, and the graphemic, since compu-
ters haven't yet learned how to talk. LRS not only looks good
on paper; at least one computer understands it. Differently,
it is an effective, if not yet efficient, system of communication
between man and machine.

In the course of time, the Linguistics Research Center
developed further interests -- or discovered further needs. One
of these was that of handling information. I'd rather not define
information. Whatever it is, one may speak of information stores;
an example is the type of entry we find in a desk dictionary;
another is the sort of response one gets from an informant to a

specific question. I haven't mastered all the relevant lore,
nor even the report on the <u>Proceedings</u> <u>of</u> <u>the</u> <u>Symposium on</u>
<u>Education</u> <u>for</u> <u>Information</u> <u>Science</u>, (Washington, Spartan, 1965)
which was held September 7-10, 1965. But some entities through
which man seems to handle information are often called concepts.
Using this term I might state that a second system -- Information
Maintenance System -- has been designed at LRC to handle concepts,
not all of which are parallel in complexity. Besides dictionary
entries and informant responses IMS will handle entire documents.
Just as LRS manages the stuff of language in various hierarchies,
so IMS will be able to handle information varying in complexity.
Without commenting on their ultimate adequacy, or economy, I
would like to note simply that it seemed appropriate to develop
two discrete systems, one for handling the mechanism man uses
for communication, the other for the ends of communication.

Obviously the two systems would be more effective if they
were interrelated. And it's almost superfluous to add that
the interrelationship has been effected, and that it is known
as an interface.

In comparing this arrangement with descriptions of human
language several questions arise:

1. Are the two systems those that should be maintained
on the basis of a thorough study, or are they ad hoc?

2. Whether ad hoc or not, what sort of interface should
be established? Should the interface be a totally separate
system; should it be incorporated within one of the fundamental
systems; or within both?

3. A third question which I do not propose to examine, for
I consider it premature, is whether the two systems should be
distinct.

My resistance to even considering this question may give
hints on my answers to the others. I view this design as
essential in the current stage of our understanding of man-
machine communication, and of computer language. In support of

this answer we may quickly review the bases of these positions with regard to morpho-phonemics, the linguistic interface which has been extensively studied.

In the nineteenth century almost all linguistic activity was concerned with morphology. Grammars of Latin, Greek, Germanic languages and others presented long sections on morphology; though phonology was included in the grammars, it was primarily intended to complement morphology. The predominance of morpho-logy may even be demonstrated by the acclaim given rare phonological insights, like those of Verner and Grassmann. But only at the end of the nineteenth century was there sufficient interest in phonetics to achieve adequate understanding of it.

As we all know, this understanding led to great concern with phonemics, and to its development in the course of this century. Phonemicists cultivated phonology with as much concern as the grammarians had devoted to morphology in the nineteenth century. In the second quarter of our century, phonology might make up the greater part of a grammar, or be an end in itself.

These two concerns led to two distinct descriptions of segments of language. Clearly the descriptions needed to be brought together. If you are a neat level man, the most appeal-ing procedure would be to leave the structures which were already erected and to set up another beside them. I do not mean to be derogatory about the achievement, or about methods adopted to lead to it. Advances have come from delimiting one's concern. But after some understanding has been achieved, such concern may be broadened and the boundaries, seams or interfaces between them reduced in dimension. Such reduction is now being vigorously proposed by one group of linguists. But in the development of control over our area, the separation of sub-areas -- with subsequent interfacing -- has been a successful programmatic and experimental procedure.

We may now ask whether this procedure should be applied further. If so, what are to be the sub-areas, what the interfaces?

The number three has long haunted western thinkers. As
a product of western culture, linguistics understandably follows
the parade. I merely call attention to this phenomenon, for
it is my charge to ask questions, not answer them. Further, it
would be ungracious, and unwise, to cite approaches of parti-
cipants of this meeting. To illustrate the preoccupation with
three, I then turn to Robert E. Longacre's "Prolegomena to
Lexical Structure," Linguistics 5 (1964) 5-24. Following Pike
and Trager in suggesting a scheme of "trimodal linguistic
structuring", Longacre sets up an axis of phonology, grammar
and lexicon; a second of particle, string, and field. While
Longacre assumes the second axis for each of the components
of grammar, he suggests that the particle has been most useful
in phonological analysis, the string in grammar, and, more
tentatively, the field in lexicon.

By Longacre's approach any text would be analyzed for the
three dimensions of phonology, morphology and lexicon. In each
dimension we may posit meaning, though the phonological will be
rudimentary. Though he is not explicit, meaning to Longacre is
apparently the interrelationship found in sets, classes or
fields. By this approach we would seek out something like Harris'
"equivalence classes." Much of the meaning of a text would be
conveyed in the lexicon, some in the grammar, little in the
phonology. Since three levels, with three units: phonemes,
morphe es, lexemes are proposed, this type of analysis would
make relatively heavy use of interfaces.

In this way it would contrast sharply with the approach
which says that "syntactic structures are the foundation on
which the rest of language (analysis) should be constructed."
This approach, by Tabory's views in a recent, apparently hastily
written article, states that "work in semantics means work in
syntax;" R. Tabory "Semantics, Generative Grammars, and
Computers," Linguistics 16 (1965) 68-85. And a bit later: "the
part of semantics treated by syntax has to be made explicit by

extensive morpheme classification." To which of these views
is this conference directed: that of a sub-syntax semantics,
or some larger domain? If a larger domain, is it co-extensive
with Longacre's?

If co-extensive with Longacre's we are presumably
concerned with a portion of the logical semantics of Tarski-
Quine. Although their point of departure is logic, not language,
their semantics consists of two sub-areas: a theory of meaning
and a theory of reference. If we include the second in our
discussion we would have to deal with "truth with respect to
an extra-linguistic situtation." If only the first, we
probably would not have to proceed beyond Harris' discourse
analysis, for under the theory of meaning our chief aim would
be to decide whether "two statements are logically equivalent
or whether a statement is logically true." Then at least our
interfaces would be reduced.

Whatever our delimitation, I should like to claim a key-
note speaker's privilege and ask that we be moderately consis-
tent in our terminology. It would be easy to cite uses of
semantics to include virtually the entire domain of language.
At a contrasting extreme, scholars displace the term with another
to stake out their area with distinctive markers. I'm happy
to see that our guide has not fallen into either pit of despera-
tion.

I also urge other aims, among them a specification of the
goal of a semantic theory. Fodor and Katz specify "the basic
fact that a semantic theory must explain is that a fluent speaker
can determine the meaning of a sentence in terms of the meanings
of its constituent lexical items." (See "The Structure of a
Semantic Theory," in The Structure of Language, by Jerry A. Fodor
and Jerrold J. Katz (Englewood Cliffs, Prentice-Hall, Inc. 1964)
pp. 479-518, p. 493. The addition of this statement to the
original version of this article in Language 39 (1963) 120-210,
as well as a variation in the citation of Tabory, suggests a

certain fluidity in semantic study of the present. R. Tabory,
"Semantics, Generative Grammars, and Computers," Linguistics
16 (1965) 68-85, cites this sentence with 'morphemes' instead
of 'lexical items' (p. 77). But why should we determine mean-
ing by sentences? Again I withhold my point of view, but I
would expect this conference to specify the entities most promising
for work in what we define as semantics.

I would also expect the conference to be clear about the
role of semantic entities in language. Again I accord Fodor
and Katz the distinction that comes to recent widely read
commentators on a subject as I object to their statements,
though I cite Tabory's commentary. According to him "lexical
meanings are the primitives of the Fodor and Katz theory:
being entirely intuitive they cannot be formalized." I find
it difficult to understand why lexical meanings are more intui-
tive than are the primitives at other levels of analysis, such
as distinctive features. We have long departed from the view
that we should operate with substance rather than form ·- in
Saussure's terms. Distinctive features, phonemes and even
morphemes are intuitive, or in other terminology they are fictions.
We will use different fictions in semantics as well. And the
fictions, or intuitive entities we select are not the focus of
of our formalization; this is rather the relationships we posit
between these entities. Formalization, though it may be more
complex than that for syntactic study, will therefore be
required in semantic study.

But when we formalize, we should be aware that we are
simply applying a means to manipulate our data. Formalization
provides great sport for the formalizers, but unless it is
relevant to the data in a particular field it adds nothing to
our knowledge. You may substantiate this statement by checking
recent formalizations of linguistic data and observing that
unformalizing predecessors managed those data capably, possibly
more capably than their formalizing successors. Yet we endorse

formalizations in the expectation that they are developing tools by which we will manage ever larger and more complex data, the data of semantics.

In managing the data of language, all kinds of linguistics, including computational linguistics, are founded on theory. But the special goal of computational linguistics is to proceed to the programmatic from the theoretical. Linguists today are undertaking a more complex task, and a larger one, than has yet been achieved in the study of language. In my view it can be undertaken only because we have devices to manage the complex masses of data. The past few years have equipped us with skills to use these data devices, skills that will increase in sophistication as the devices and their users develop. This conference should give us a focus around which we may pursue theory to practical programs. It is difficult to see why it will not contribute toward a break in the wall which has hitherto surrounded semantic theory.

I.

THE OUTLOOK FOR COMPUTATIONAL SEMANTICS

Yehoshua Bar-Hillel
Hebrew University
Jerusalem, Israel

Let me first apologize for the utterly inadequate title
of my talk. At the time when I submitted the title, I
thought I would be able to get a full conception of the whole
field of computational semantics sufficient for a talk about
the outlook for the whole field. In the meantime something
happened to me that, had I been wise enough, I should have
predicted would happen; namely, that the more I got involved
and the more I was thinking about it, the more the field as
a totality started to recede, and the more innumerable details
began to come up to the front, and it is obviously pointless
to attempt to predict the future of a whole new field in twenty
or twenty-five minutes. So I am afraid I will have to do
something much less than what my title might have promised to
you, and perhaps it is much better so.

Let me start, appropriately, with a couple of semantic
remarks to the phrase "Computational Semantics" occurring in
the title. I can predict, almost with certainty, that this
combination of two very fashionable terms, "semantics" and
"computational", will soon become itself so fashionable so
that it will be jumped upon from various sides and will quickly
become as ambiguous, maybe more so, as each of these terms is
separately. Particularly I think at least three meanings of
this term are already in the offing (and may have already
showed up in last night's informal discussion).

The one meaning is "semantics of computer languages".
I think this is a highly interesting field. I have dealt with

it on other occasions, but for lack of time shall not do so today.

A second meaning which the term already has or will have is that of using computers as an aid for producing semantic theories of natural languages. Here again I wish I had more time and could argue my view at length. Since I do not have this much time, let me state quite dogmatically that I do not think, contrary to what other people are already attempting to, that computers could possibly be of any serious help for the mentioned aim, i.e. they could not do much beyond supplying statistics and concordances and things like that.

Let me then turn to the third meaning, which I believe is still the most frequent one; namely, of using computers for analyzing the semantic structure of sentences, in some natural language, English or Russian or what have you, in such a way that the output of this analysis will in some way or other more clearly, more precisely, or more overtly, exhibit the semantic structure or structures of these sentences.

The first questions that have to be answered are "Why do so altogether?" "Who is interested in this job?" "Why should we want to input an English sentence and output something that will exhibit the semantic structure of this sentence more precisely than it was to begin with?"

Well it seems that one aim of this job is translation. It now seems that for the purpose of computer-aided translation the semantic structure of the sentences to be translated has to be exhibited. Without such exhibition of structure it is not very likely that an adequate computer-aided translation will be forthcoming.

Another use of semantic analysis is information processing. It seems to be almost generally agreed at the moment that with natural language input as such, without preliminary semantic processing -- for which I shall use here

the term "standardization", and have been using on other
occasions the term "sterilization," -- one cannot, certainly
not at the moment, maybe not even in the foreseeable future,
do much about processing this input for the innumerably many
purposes for which this input could be brought to use. But
in order to arrive at that standardization, it seems that
going over the meaning or meanings of the input is of par-
ticular importance.

Something else. A minor side effect of computational
semantics would be to exhibit hidden ambiguities, and on
occasion a computer might do this better than human beings.
This has been often put to a psychological test, and one
has found that human beings, when 'n appropriate conditions,
very often understand a given utterance in one particular
way, which is indeed one of its meanings, but is only one
of its many meanings, even within the whole context.

Just recently Martin Joos told me that on a certain
occasion he uttered a request which half of the people around
understood in one way and half in another way, while nobody
was aware that his request was ambiguous. In such special
cases, an appropriately programmed computer could more easily
come up and say, "Well, these are the two meanings. Now pick
whatever is appropriate."

One can also envisage that one could want to have a
computer test for consistency or any of the many other logical
relationships between the input statements. However, I hope
that you are all aware of the fact that for medical diagnosis,
for jurisprudential purposes, and presumably even for straight-
forward scientific purposes, so long as the input is given in
some natural language and not in some formalized language, these
tests cannot be performed by purely syntactical means. The
inconsistencies, if there are any, will in general only turn
up through what is called meaning analysis or semantic analysis.

Obviously if semantic analysis of natural language texts could be done with the help of computers it would be a major achievement.

In the rest of my lecture -- which so far was pure description -- I intend to make only two points. In the discussion, if we have time, other things might be brought up.

My first point is the following: It is my belief that the existent semantic theories of natural languages, including those that were proposed during the last two years or so, are woefully inadequate and that something very central has been missed.

Just for the sake of illustration let me refer to the Katz-Fodor theory since this theory is presumably best known to the participants of our meeting. But what I am saying now should apply to any other semantic theory.

The major cause of the inadequateness of the Katz-Fodor theory lies in its conception of a semantic theory being composed of a dictionary and projection rules. The dictionaries that they have in mind differ from standard dictionaries but not to a degree that will affect my remarks.

You might want to find out for yourselves why dictionaries should have obtained such a prominent role in the thinking of the people in the field. But whatever the reasons, I have a strong conviction that to state the meaning rules, or semantic rules, or whatever other term one is going to use in the future for this purpose, in the form of dictionary plus projection rules is just not adequate at all. The meaning relationships that have to be described in these rules cannot be described in those two forms alone.

I would not want to say for a minute that these are not also forms in which to render the meaning relationships. Of course they are. I don't want to abolish them. But they are not enough.

The clearest discussion of meaning relationships though originally related mostly to formalized languages, are due

to Rudolf Carnap. His term for what we have come to call "meaning rules" is "meaning postulates," again because he is thinking mostly in terms of constructed languages, so that for him those rules are postulates, whereas for us those rules are empirical findings.

The meaning rules, the rules that describe the meanings of terms and phrases of natural languages, cannot be handled by dictionaries alone. The meaning connections that hold between the various terms in natural languages, cannot be handled by dictionaries, extant or foreseen, alone. They are unable, in principle, by their very form, to take care of all the complex meaning relationships.

Let me present only a trivial example at the moment. There are infinitely many others. By virtue of the meaning of the English expression "is warmer than," if A is warmer than B and B is warmer than C, then A is warmer than C. This is a fact of English meaning. It is not a fact of logic. Anybody who understands the meaning of "is warmer than" must consent that the relation denoted by this expression is transitive, to use the logical lingo.

Now, of course, nothing of this kind could possibly be treated by a dictionary. Where will you find in a dictionary of either classical or the Katz-Fodor type an entry for "is warmer than"? You have an entry for "warm", of course. But this entry could not possibly take care of the transitivity of "warmer than". Nor can the projection rules account for this extremely simple fact and innumerable others.

The meaning rules that in combination will create a semantic theory will have many different forms - I don't know how many. One might want to classify these rules and see how many of them can be handled by something like a dictionary. It is, in general, advantageous to replace algorithms by table look-up. I therefore hope that even in the future, dictionaries will be able to carry a good amount

of the load involved. But they will not be able to carry the whole load.

This brings up the second point. Due again to certain highly interesting historical developments which I shall not try to sketch here, linguistics has become divorced from logic for most serious linguists, in particular for most American linguists.

The result was extremely unfortunate. This divorce between logic and linguistics is intolerable. It is inherently a wrong view.

As an example let me come back to what I said a few minutes ago. Most linguists, presumably most of the linguists sitting here, would say that it is not the business of linguistics to state that the relation "is warmer than" is transitive. (They might not even understand this way of speaking.) Without using this "logical" terminology, they might insist that it is not the business of linguistics to interfere with whether one is entitled to deduce from the facts that A is warmer than B and B is warmer than C that A is warmer than C.

But this looks to me utterly wrong. Obviously it is <u>only up to the linguist</u> to tell, to explain, to exhibit, to clarify the meaning of "warmer than" -- and uncountably many other phrases in English -- in order to enable anybody to deduce from these two premises the conclusion.

A logician as such, of course, will not take this task upon himself, because the straight logician will say that his profession has nothing to do with the English language. What is happening in the English language is not his business. What is his business is to state that if a particular relation is transitive, then such and such. "If A stands in the relation R to B, and B stands in the relation R to C and R is transitive, then A stands in the relation R to C." But whether the expression "warmer than" is transitive, what can he, qua logician, say to that? He is not, <u>qua</u> logician, an expert in the English language.

Let me repeat: It is the business of the English linguists, and of them alone, to provide the information that entitles anybody to draw the mentioned inference.

In general I would say that there has been, in connection with this dictionary business, an incredible overestimate of the role of synonymy and paraphrasability in all linguistics, but strangely enough in particular in modern linguistics. The terms "synonymy" and "paraphrasability" -- as well as some of their variants -- have become the most basic terms for modern semanticists. This is again historically understandable, but still essentially a very strange development because, as a little logic and perhaps even a little common sense will tell you, from such symmetrical relationships, and both of these terms denote symmetrical relationships, it is either impossible or in any case very hard to define certain asymmetrical relationships which definitely are of extremely great importance in semantical thinking. Such notions like "hyponymy", or "meaning inclusion" -- the property expression 'A' is hyponymous to 'B' if and only if anything that has property A also has property B but not vice versa - clearly cannot be defined by synonymy, though it is clearly possible to define synonymy by hyponymy.

But the fact that linguists think that paraphrasability and synonymy is their business, while hyponymy is not and belongs to logic, because it lies at the basis of inference and drawing conclusions, is a strange development which has been quite fatal to modern linguistics, and particularly to modern semantics.

As one conclusion from these considerations, I think that light can be shed on the question of the borderline between semantics and syntactics, a question which has already been discussed and will probably come up many times more during our present meeting. I presume you know that the M.I.T. School has been changing its mind every few months on this quite confusing question.

As soon as we understand that dictionary-type rules
or rules of paraphrase are <u>only</u> <u>part</u> of the totality of
semantic rules, then the question of the status of, to illus-
trate by one of the standard examples "Misery loves company",
whether this sentence is syntactically acceptable, but semanti-
cally somehow not quite to the top of the ladder of meaning-
fulness, can be seen in a new light.

When we are asking ourselves, what is the meaning of
"Misery loves company", we cannot turn to dictionaries and
projection rules to find the answer. It is not inconceivable
that the actual meaning rules for expressions of the form
"A loves B" would be such as to assign a certain meaning to
such expressions in case A is human, but leave it without
any specific meaning when A is non-human.

The meaning of "A loves B" is in this particular case not
established by those rules which, however, should not be under-
stood to mean that the expression is meaningless. It only
means that this expression is <u>so</u> <u>far</u> <u>without</u> <u>meaning</u>; that the
existing meaning rules just are not sufficient to give to this
expression any specific meaning. This is quite different from
saying that <u>it</u> <u>is</u> <u>meaningless</u>, because if it is so far without
meaning, we can add new rules to the meaning rules of this
particular language at that particular stage, without changing
any of the old meaning rules, something that couldn't happen
for dictionary-type meaning rules.

We should realize that there is nothing wrong with having
in a language expressions whose meaning is, at a certain stage
or even at any stage, not completely determined, which will be
intelligible in some contexts but meaning-indeterminate
(rather than void-of-meaning) in others.

It might turn out to be that with regard to certain
expressions, particularly with regard to the so-called <u>theore-</u>
<u>tical</u> <u>expressions</u>, any attempt of expressing their meaning
by a single entry in a dictionary is in principle utterly wrong.

We already know that theoretical expressions get their mean-
ing in an entirely different way. Their meaning is theory-
dependent and can only be determined by taking into account
the whole set of postulates of that particular theory. But
the issue is too complicated and technical for us to discuss
it here. My final conclusion is that inasmuch as semantics
is concerned, we have been living in a fool's paradise until
this date. We knew that semantics is difficult. But we
kept fooling ourselves to believe that we at least knew the
type of semantical rules that would be employed, so that our
only problem was to get sufficient empirical information to
be able to state all our semantical findings in the form of
a dictionary plus projection rules.

We must now realize that this was illusion. We will
have to live up to the fact that semantic rules in general
will be of many additional types. It seems to me that for
the time being we should let them have every form that seems
appropriate for a problem at hand and that only much later
should we start again and see whether these innumerably many
ways of forming semantic rules can be reduced to a more
manageable subset. Some of them will turn out to be rules of
paraphrasability and projection. Others, of course will not.
Only when this is accomplished - and I would not dare estimate
today how much time this will take, - will it become feasible
to develop computational semantics, in the third meaning of
this expression, i.e. to determine with the help of a computer
the meaning or meanings of any given natural language text.
If this estimate will be regarded, as presumably it will be,
as another expression of my now well-known "pessimism", I am
afraid it can't be helped. My own way of putting it has
always been that I have had the misfortune of arriving at
realistic evaluations quicker than most other workers in the
fields tended to do, so that it has been my unfortunate

privilege to insist from time to time that other people's
thinking is marred by a good amount of wishful thinking.
As I see it, I am not pessimistic, I am realistic.

DISCUSSION

RUBENSTEIN: I have no argument with your general evaluation
of semantics. Indeed, it is far beyond my competence to argue
against that. What I would like is to express my reservation
about what you implied by way of scientific procedure.

My connection with computers is very marginal. What
I like about the computer is that somehow it enables me to
set a sub-goal. And possibly it enables me to evaluate how
closely I have come to the goal that I set out to achieve.

In short, what I am simply saying is that I feel we have
to set up sub-goals, but with this notion: We should try, as
far as our intelligence, our foresight, enables us to, to try
and set up sub-goals such that what we find when we have
achieved one goal can be added onto our next sub-goal.

BAR-HILLEL: I don't think I could possibly have any quarrel
with that. I also think that this is exactly what has occurred.
You see, the most important sub-goal that semanticists, by
that name or any other, have set themselves, was the estab-
lishment of the equality of meaning, or the simple term synonymy.
All right; no quarrel -- except that in the minds of many
semanticists this particular sub-goal has beclouded the issue,
as for obvious psychological reasons on occasion happens.
They have gotten the feeling that this is now approximately
all there is to this, and I think it is important that they
should realize that this is an important but still a quite
moderate sub-goal, and when they are through with that, and
heaven knows how many years or centuries this will take, it's
a far cry from this to total semantics of natural languages.

YNGVE: In general I agree with what you said. I may disagree
on some particular point. But what I would like to do is to
ask two questions more from the point of view of trying to

get some clarification of what you said. The two points are the following:

Take the first point. You talked about the need for information processing by computers of some type of standardization. I think I understood you when you said "standardization" but I am not sure exactly what you meant by this standardization, and I wondered if you could elaborate a little bit on that. That is my first question.

The second one is, in the beginning you gave three ways in which we might talk about semantics. You dismissed the semantics of computer languages. I would agree with your first point.

The second and third was using computers in two different ways, and your second point was something that you believed there was no chance that computers could be of help in; and your third point embraced a number of areas where you thought computers would be of help.

Now, my question is to ask for some elucidation on your second point. Precisely what is it that you think we can't do with computers, because maybe I disagree with this. I don't know. I haven't understood it.

BAR-HILLEL: Yes. I think it will be quite generally agreed upon that for the time being, unless something radically new develops of which I think nobody at the moment foresees what it can possibly be, a direct operation upon natural language texts is out of the question, at least to any serious degree. So before you are going to do something, if you are going to process it through a computer -- use any term you like, standarization may be quite all right -- this natural language text has to be standardized by non-computers. This is obvious because, among other things, innumerably many other things, in natural language use so-called "X-er" expressions, hundreds of expressions whose exact meaning within a given sentence is

completely dependent on both the linguistic and non-linguistic
context around it, surrounding it; both on what other utterance
of language has preceded it, and in some cases will come after
it; and more importantly on who is speaking to whom, when, and
under what conditions, and so on. And all these things for the
time being, obviously, are utterly beyond the ability of any
computer. So that the rephrasing of the natural language texts
into a way in which a computer could possibly do anything
at all has to be done in other ways.

As for computers, let me again say that many people have
thought that one can use computers in order to develop theories
of natural languages, to write syntaxes, semantics, I think.
Maybe a number of people who are sitting here are even involved
in that. So, one may use computers in order to arrive at
theories of language.

Here again I can't elaborate, but my only serious argument
is that it is for me just utterly inconceivable; not that any-
body has ever said anything about it, but it is utterly incon-
ceivable how such a thing can be done, both in principle and,
more so, in this particular case. Theory elaboration, to come
up with, say, grammar, not to say the semantics of a natural
language, is often used in the sense that this is a theory of
natural language. Theory construction is something that at
the moment is by many orders of magnitude out of the reach of
computers. I can only deplore that a number of people in the
United States, and particularly also elsewhere, have started
even talking about this as a serious subject.

So I think we should be, in this case, utterly realistic
and realize that this is something of which it doesn't even
pay to seriously think at the moment. So that the use of
computers in order to arrive at linguistic theories seems to
me, or the outlook would be, out of the question.

However I think, since you asked the question, that one can use computers on occasion to test. After you have, by whatever means, by using your intelligence, come up with a linguistic theory, under very extreme cases you might be in a position to use a computer in order to text certain things. So if you have arrived, by whatever means, at certain grammatical rules to which you at the moment cannot see whether there are any exceptions or not, you might want to run these things in certain ways through a computer, and let the computer, following those grammatical rules, generate all kinds of things. Then you might look at them and say, "Obviously there is something wrong with my grammatical rules."

So, to use the computers for those purposes, obviously I wouldn't have the slightest objection, but there is an enormous step from this to the belief, which I am afraid some people share, that one can use computers and, with the help of computers, arrive at theories at all, and in particular at linguistic theories.

YNGVE: I agree, now that I understand what you mean.

II.

SOME TASKS FOR SEMANTICS

Uriel Weinreich
Center for Advanced
Study in Behavioral
Sciences

When Harry Josselson first asked me to come to this
conference on computer-related semantic research I thought
it was a case of mistaken identity, because I have never
done any computational work myself. But I did welcome the
opportunity of coming here to learn. Like many other lin-
guists, I am aware that over the past decade or more, the
considerable frustrations and failures of computational
linguistics have been productive of new linguistic insights.
As in other fields, the failures of technology have been
greater boosts to the progress of science than technology's
successes.

My own work in this area - more of an armchair nature -
is very preliminary and very programmatic,[*] and this is
reflected in the title of this short informal talk, "Some
Tasks for Semantics."

I have been concerned in particular with outlining
a type of semantic theory which would be compatible with
a generative approach to syntax and which would give us
some guidance about the way we should speak about the se-
mantic form of complex expressions, expressions of a com-
plexity up to the degree of the sentence.

[*]For additional detail, see my "Explorations in Semantic
Theory," _Current Trends in Linguistics_, Vol. III, ed.
T.A. Sebeok, The Hague, 1966, pp. 395-477.

In the past, most semantic work done by linguists has been
concerned with individual items or with items in paradigmatic
relation with each other, rather than with the combination of
items in a sequential (or still more complex syntagmatic) order.
It is in this area of combinatory semantics, I think, that
some new formulations are badly needed.

Looking around to neighboring fields, a linguist finds
two possible models which he might consider using. One is
offered by an associational psychology. According to it, when
two simplex expressions, the meanings of which are given, are
combined syntactically, there results an association between the
meanings of the components. One implies the other, and so on
throughout the chain.

I don't think we need spend much time in showing why this
would not be adequate for the semantic explanation of an
arbitrary sentence. Of course we might say that in a sentence
like The tablecloth is white there is an association formed
between the meanings of tablecloth and white, but I don't know
what sense it would make to say that there is also an association
between the meanings of the and tablecloth; and, after all, we
are accountable for that, too.

There are many other typical occurrences which simply
could not be dealt with in terms of associations between elements
in sequence. For example, a construction might end between two
elements. In The girls left there is some kind of association
between the meaning of girls and left. But if we should say
The men who helped the girls left, no such association takes
place. Because of the well-known hierarchical structure of
discourse, a simple associational account would simply not do.

The other model available to linguists from an adjacent
discipline is a Boolean-algebra model which logicians are very
familiar with. Its application would amount roughly to this:
if we have two expressions, and the meaning of each is stated

in terms of some semantic features of that expression, then,
when these two simplex expressions are combined, there takes
place an addition of the features. For example, if we say
white tablecloth, the expression contains the semantic features
both of white (things) and of tablecloth. This addition of
features or intersection of classes is something that is
familiar even to non-logicians. It is something that has
been tried in linguistics on various occasions.

But this model too, I think, is quite inadequate, although
the reasons are perhaps not so obvious. An account like this
might be possible for very simple predicate sentences like
The tablecloth is white or The girl is tall -- at any rate,
for some parts of those sentences. But if we take something
like The girl laughed infectiously, for example, we cannot
possibly say that there is an "addition of the features" of
girl and laughed and infectious(ly) (I leave out a lot of other
formatives in a sentence like that). We are, in that sentence,
clearly not postulating any entity like an infectious girl.
There seems to be one predication, or one addition of features,
between girl and laugh. A laughing girl is indeed postulated.
And another addition takes place between laughing and infectious.
(Her laugh was infectious is another way of paraphrasing the
same sentence.) But there is no overall addition of the features
of these three content elements of the sentence. It is as if
we had a two-dimensional structure: two predications "at right
angles" to each other. In this type of sentence there just is
no predication in a single plane.

When we come to transitive expressions, again the model
of adding features or intersecting classes does not work. The
girl ate an apple: there is simply no semantic entity created
through that sentence which belongs both to the class eating
and the class apple.

In other words, it would seem that in an arbitrary sentence there are syntactic nodes at which a semantic process takes place describable in terms of feature-addition or class-intersection; but there are also many other nodes in the structure of a sentence where no such process takes place. The nodes that fail to produce "semantic linking" are of several types:

(a) Modifiers "in another dimension," e.g., manner adverbials in relation to verbs.

(b) Transitive constructions, e.g., between verbs and their objects, or between prepositions and their objects.

(c) Elements entering into a sentence for quantification purposes (including, perhaps, the whole determiner machinery of a language), e.g., the relation of the to girl in the girl.

(d) There also seem to be "modalizing" elements in a sentence whose function is to qualify or to restrict the way in which something is linked. In The girl seems happy, seems appears to qualify or limit the kind of linkings between meanings of girl and happy which would otherwise take place.

This account, I confess, may sound disappointing because it turns up so much non-linking (non-predicative) structure in sentences. Predicative structures and their transformational derivatives are far more attractive. The whole history of logic attests that when you have expressions analyzable into subject-predicate form you can calculate with them, you can make inferences, you can construct syllogisms and prove theorems. On the other hand, linguistically transitive expressions typify the impossibility of calculation. For example, from John loves Mary and Mary loves Tom it does not follow that John loves Tom. That is to say, the theory in which transitive expressions function is of extremely limited power.

To be sure, some linguistically transitive expressions happen to be logically transitive also. If we say <u>The glass contains water and the water contains a mineral</u>, we might infer correctly that the glass contains a mineral, although there are some problems there, too. But this, as I say, is a special case. It is not in general true that what is linguistically transitive is also logically transitive. And certainly there is little semantic work which all the quantificational machinery and the modalization machinery can do for us, in contrast to the predicative relation.

So I realize, and admit, that to take a syntactic analysis of a sentence and to say that there are few nodes in this structure where semantic linking takes place, while at all the other nodes the semantic process in effect is not of the linking type, is a frustrating and a negative finding. It is important nonetheless; in fact, the failure to realize it is one of the main weaknesses of the Katz-Fodor theory.

If you actually put that theory to work, you come to the result that, say, <u>Cats chase mice</u> and <u>Mice chase cats</u> have the same meaning: the two sentences contain the same ingredients and the semantic process obliterates the syntactic difference. Yet obviously we would prefer an account which would show why their meanings <u>are</u> different.

Logicians may want to object that the subject-predicate logic which I find insufficient as a model of sentence semantics has been superseded by the far more flexible logic of relation. Instead of having to say that <u>John loves Mary</u> is of the same structure as <u>John is tall</u>, we could utilize the logic of relations in such a way as to say that in the former sentence the terms <u>John</u> and <u>Mary</u> are both arguments of a particular relation -- <u>loves</u>. No doubt this relational formulation does account for many more types of expression than a subject-predicate analysis. But it is a case of throwing out the baby with the bath water,

for it fails to show that in a predicate of more than one place,
of the the arguments remains basically in a subject relation to
the relation (predicate) term. That is, even when we have <u>John</u>
and <u>Mary</u> agruments, let us say, bound by a certain relation --
<u>love</u> -- the terms <u>John</u> and <u>love</u> are in a subject-predicate
relation nevertheless. Of the two arguments, <u>John</u> in this
formula is still in a privileged or special place. If there
are any generalizations to be made about people who love, let
us say, we can utilize this relation for purposes of inference:
e.g., if all lovers are happy, then John is happy. We could,
in a general way, put to work the pair consisting of the relation
term and <u>one</u> of the arguments, but we could not do this to all
the others.

I have talked about the semantic processes that should be
looked for in the structure of a complex expression, and have
argued that there is a kind of irreducible structure, and that
it would be incorrect to say that in the semantic interpretation
everything becomes linked in the long run. But what about the
interrelations of simultaneous semantic features, in the meaning
of a component expression, let us say <u>girl</u> or <u>tablecloth</u>? It
is generally assumed, perhaps merely for the sake of argument,
that these component semantic features form an unordered set;
that is, the semantic features constituting the meaning of a
component term (a lexical entry in a dictionary, let us say) form
an unordered set. Indeed, I find that the references to feature
ordering in the Katz-Fodor account are vacuous; they are not
justified and are not put to work in the theory.

If we think of the semantic features of a component
expression as somehow reconstructing the dictionary definition
of that expression, and if we see that the dictionary definition
is itself a sentence in a language, subject to the same kind of
non-linking semantic processes as our original object sentence,
then it is clear that there is a syntax of the simultaneous

features of a component expression as well. In fact, I want
to argue that it is in principle the same kind of syntax as
you have in the sentence whose analysis we began with.

If, for an expression like girl, we want to invoke the
simultaneous component features 'young' and 'female', then
these components are indeed in a linking relation, and therefore,
if girl appears in the predicate of our object sentence, and
the predicate links with the subject (e.g., Our guide is a
girl), then all the linking elements in the definition of girl
will be linked with the subject of the sentence: we infer that
our guide is a female and is young.

But if we have a transitive relationship within a definition
(for example, in "A chair is something one sits on," the relation
of sitting to chair is transitive rather than predicative), then
there will be no linking. If X is a chair, we will not conclude
from that that X is sitting. On the contrary, X is sat upon.

When logicians talk about semantics as a domain of research,
they assume that there is an object language distinct from the
metalanguage of semantic description (which has rules of its
own). But when natural languages are used as the tools of
their own semantic description, there is a complete continuity
between the expressions in the object language and the "semantic
rules," which are also statements in the object language. They
are statements with a special function, but syntactically they
lend themselves to the same kind of analysis.

And the last point, which is related to this, is a plea to
linguists and to other semanticists to cast off the shackles of
a dilemma which has been inherited by linguistics from logic;
namely, the dichotomy of expressions into the well-formed and
the uninterpretable. I think all the attempts to construct a
semantic theory compatible with generative grammar have remained
in the grip of this unfortunate dilemma. At best, previous
attempts have given an account of what is well-formed, and for

that which is not, they have tried to say in what way it is
deviant, but not to go one step further and say exactly what
it means.

The consequence of accepting this dilemma is that if you
want to have a semantic theory which accounts for all sorts of
deviant uses of language (and I think they are just as legitimate
and frequent as non-deviant ones), you will have to have infinite
dictionaries because you will want to foresee all the possible
misuses of the word which every speaker will nevertheless under-
stand. What is needed instead, I think, is a semantic theory
by which meanings can result from the combination of elements
which were not stored to begin with in the dictionary. To
take Mr. Bar-Hillel's familiar example, if the word _love_ by our
account requires a human subject, and if we then use some noun
which doesn't have the feature 'human' in it, what I would expect
the semantic theory to do is to show how, by being so used,
the noun has the feature 'human' _imposed_ _on_ it by the verb.

This means that any account in which the features of verbs
are merely selectional, and have no power of imposing themselves
on the noun material either side, is simply not capable of
dealing with such uses which, though deviant, are nevertheless
completely transparent and semantically effective.

DISCUSSION

SPARCK JONES: I wanted to say that Dr. Weinreich said he
hasn't come across any work in which syntactical structures
feature definition work. I should say that at the Language
Unit we have been doing this since at least 1958. We came
precisely to the conclusion that associative combinations of
features were no good, so we tried a very simple structured
system. It also has the feature that you were wanting, that
you continue with the same kind of structure from a unit like
a word up to the sentence.

WEINREICH: It is to learn things like this that I came here.

ROSS: I would like to comment on one example of yours, about
"The girl laughed infectiously." You assert that there is no
predication or association of "girl" and "infectious." I
would take issue with that, because if you note the non-
existence of sentences like "The tree fell down infectiously"
there is a whole class of what have been called "manner adverbs"
which depend on features of the subject. Whether this is syn-
tactic semantics, I suspect it is really the underlying form of
"The girl laughed infectiously," which would be something like
"The girl was infectious in laughing."

   So I would argue that there is a syntactical relationship
on one level between "girl" and "infectious."

   The other thing is, you bring up this problem about the
asymmetry of the treatment, the problem of transitive verbs,
may I say, as opposed to adjectives. I myself am opposed to
any treatment of adjectives and verbs which doesn't treat both
the same. I think Katz and Fodor originally drew the wrong
conclusion, treating them all as simple linking or association
or whatever you want to call it. I think the opposite conclusion
is correct, because there are many transitive adjectives, like

"proud of", "mad at", "helpful to." Clearly "John is proud
of Mary" is different from "Mary is proud of John" in pre-
cisely the same way that "John loves Mary" and "Mary loves
John" are different.

In logic there is no difference except in the number
of arguments between "John is tall" and "John loves Mary."
So I would argue for a more unified relational treatment of
all kinds of predicates.

WEINREICH: But if you have a two-place predicate, wouldn't
you want to say that one of the arguments is in a special
matter relation to the relation term?

ROSS: I see what you mean, and I think that maybe what was
the cause of being led down the garden path earlier is that
there is some very difficult sense which is philosophically
unclear, and link-wise particularly unclear, with the word
"about." In some way "John loves Mary" is about something
different that "John is loved by Mary." There has been some
work on this in Czechoslovakia and a little in this country.
It's really extraordinarily poorly understood.

I don't know if this will be derivable automatically.
I mean, however one should try to capture this notion of
"aboutness" and topicality or something like that; maybe it
will be automatically a function of, presumably of, the derived
structure, that the first noun phrase in a derived structure
is the topic of the sentence. I don't know.

WEINREICH: I was thinking of something a little different.
I was thinking of the fact that many transitive expressions,
adjectives like "proud of", or verbs, convert very easily
into intransitive ones and there is very little of a gap felt.
And we can see these as either having a further place in the
predicate or not, and it is this optionality of the further

place, that we can have unspecified subjects. We need a lot
more machinery for that, rather than have unspecified objects.

ULLMANN: I agree very much with what you (Weinreich) said
but I am a bit puzzled about what you said on deviant features.
I think I am not misquoting you when you said the deviant
features are as frequent and as legitimate as non-deviant ones.
There seems to be some contradiction in terms here. Isn't
deviation itself a statistical concept? And at what point would
you exclude the obviously or idiosyncratically deviant from
your generative grammar? I am thinking of a recent article
by Jim Thorne, of Edinburgh, in the new Journal of Linguistics,
where he takes an example from Cummings. I think the line is
"He danced; he is dead."
 He said, "There are two alternatives. Either write all
of these very syncretically deviant uses somehow into the
rules of generative grammar of English, and thus you will see
many of their oddities, or rather have a separate generative
grammar for Cummings." This is a very extreme case, but I
should like to know at what point you would draw the line.

WEINREICH: I think that people usually resort to examples,
to quotations from Cummings, and this is too extreme an example.
It is really very much of a borderline case. I think we can
discuss some much simpler things.
 I don't have any general guide for drawing the line between
the deviant and the non-deviant, but I would want to put it this
way: Let's assume that we could agree on what is deviant and
what is non-deviant. We could also agree on what the deviant
expressions mean, and I would want a semantic account of that
from a finite dictionary, to tell us how we know what these
deviant expressions mean.
 I think that, for example, using a non-animate subject with
a verb that,according to our dictionaries and according to the

explicit meaning statement, requires animate subjects, imposes
an animate feature on the subject. We would account for its
deviancy, for the reaction of deviancy, specifically in the
contrast between the dictionary-supplied feature "animate"
and the sentence-supplied feature "non-animate."

III.

STRUCTURAL SEMANTICS: THEORY OF
SENTENTIAL MEANING

Elinor Charney

Massachusetts Institute of Technology

The well-known physicist, Pasqual Jordan, remarked recently
that in the history of the natural sciences, the solution of a
great problem often began with the astonishment about a fact
which had previously not caused any astonishment and therefore
had not been recognized as a problem at all. This remark seems
very apropros when we now consider the knotty semantic problems
that one by one have come to our attention since those beginning
efforts to translate mechanically had been put so optimistically
into operation. Before that time the ability of the human to
communicate information through the direct medium of a natural
language was thought in general to be just a minor accomplish-
ment. This ability of the human intellect was seen as a common-
place fact, so taken for granted that there were many who origi-
nally believed that all one had to do to be successful was to
teach the machine how to use a gigantic dictionary in the same
way a human did. It seemed so obvious to us that we understood
exactly how we understand the messages conveyed through language:
we first understood the "meanings" of the individual words and
phrases composing the text, and then, seeing them juxtaposed
according to various fairly simple grammatical rules plus a few
logical rules, we were able to understand the intented meanings
expressed by the whole text. The astonishment arose when those
first brave attempts at translation made it woefully clear to
us us that we did not understand at all how human beings are
able to achieve the apparent miracle of effective linguistic

communication. The very aim of translating by mechanical
methods had made a demand for a descriptive explicitness never
before demanded of any theory of grammar, much less of any
theory of meaning comprehension. This unforeseen demand forced
us to realize that whatever it was the human mind achieved so
effortlessly when he comprehended the information conveyed
through a written text, what he did could not be stated
explicitly enough to instruct the machine how to recognize the
intended meaning of the input text, much less how to translate
it correctly. Research into meaning recognition then became of
paramount concern in the fields of mechanical translation and
information retrieval; we had been made painfully aware that
unless some kind of useful solution to this problem could be
devised, there would be no possibility of success at all.

It is obvious that it is useless to ask the human being,
qua language user, to explain what the meaning of an utterance
is. He doesn't know what the meaning is; as answer, he can only
paraphrase the meaning of the original utterance by using
different linguistic techniques belonging to the same language.
And what is a paraphrase but an intralingual translation of
the meaning already comprehended! Thus, how the human being
comprehends that two utterances are synonymous is also not
understood, because the language user translates in an equally
unknown way as he comprehends the original.

To underline the complexities facing the semanticist, it
should also be pointed out that if a semantic theory is to be
adequate for our purposes, we must be able to state explicitly
how the information accumulated during a "left-to-right"
chronological progression through the linearly ordered input
text is understood; that is, our final goal is to be able to
to deal effectively with connected discourse. For practical
reasons the theoretical semanticist cannot attack the problem
of connected discourse in one fell swoop. He has to divide

the text into manageable semantic subunits. However, from the
viewpoint of the semantics of discourse, there are no pre-
determined boundaries limiting the individual information-
bearing expressions composing a text, since in a given expression
reference to information contained in previous expressions can
jump any boundary, however chosen. Moreover, to compound the
difficulties, the specified semantic interpretation of the
individual message content of each of the various expressions -
however chosen - most frequently depends upon the positional
relationship of each expression to that of others of a group
of expressions; therefore there almost always occur very
important meaning changes in the majority of expressions when
there are changes in their individual environments in the text.
Thus if we are to decompose connected discourse successfully
into analyzable semantic subunits, we have to decompose in
such a way as to be able to recompose discourse accurately
again from those same subunits, chosen as the elementary or
primary expressions composing connected discourse, without
destroying irretrievably the specific meaning the expression
may later take on when seen in the context of any meaningful
discourse. In other words, we must be able to justify, from
the point of view of the final aim of a satisfactory theory
of meaning, taking any expression out of discourse context and
analyzing it semantically first as an isolated unit. It is
possible to give a satisfactory justification if we can demon-
strate, and only if we can demonstrate, that the expression to
be thus necessarily isolated is the smallest unit of discourse
capable of expressing an individual recognizable message, and
that this meaning conveyed by the expression is constant or
unvarying when isolated, and that the specific meaning it takes
on in context is a function only of this underlying constant
meaning plus its discourse position. The only type of linguistic
expression that can be shown to satisfy these three important
requirements is the expression familiarly-known as the sentence.

For the purpose of easier and clearer exposition, let us assume temporarily that we all agree what a sentence is, and where its boundaries are to be set. Then an important part of the required justification can be given by the introduction of a distinction drawn between the concept of the <u>linguistic meaning</u> of a sentence and the concept of its <u>cognitive meaning</u>. The <u>linguistic</u> meaning of a sentence is defined as that independent meaning which is immediately perceived by the language user who has mastered the various laws governing the construction of sentences. The language user, it can be shown, has to know these sentential construction laws before he can construct the most rudimentary sentence to be used as a link in connected discourse. For example, the linguistic meaning of the English declarative sentence: <u>He robbed the store</u>, must be understood before the language user can use a <u>token</u> (i.e. an actual event of uttering a sentence) of that sentence successfully in any way. He has to know how the relative pronoun "he" functions, that "robbed" is the grammatically correct form of the verb to use when he wishes to state that the time of the occurrence of the action took place before the specific time that he actually utters the token, that "declarative mood" is the proper grammatical form to use when he wishes to express an assertion that the sentence is true, that the constituent arrangement of the symbols of the sequence specifies correctly the objective relationship in the extra-linguistic reality so that whomever "he" refers to when a token is purposefully uttered in a specific utterance act, it is understood directly that this objective referent of "he" did the robbing of the store and that it was not the store who robbed him, and so on. The additional cognitive information as to who-exactly is the <u>objective referent</u> of "he", which specific "store" and what time the event took place in reality, is given to the language user only when a token of the sentence is purposefully used, either in a specific conversation which

takes place in a unique time and space or when the token
occurs as an individual linking unit in connected discourse.

Strictly speaking, sentences themselves do not appear
in actual discourse. The notion of an isolated sentence is
a theoretic abstraction; it is a useful concept needed by
the linguist for purposes of semantic and syntactic analysis
of language. A sentence is an abstract set of sentence-tokens;
sentence-tokens alone exist as physical events. Therefore,
only sentence-tokens can appear in actual connected discourse
because the so-called individual sentences in a text are each
used once and once only in a unique position with respect to
other sentence tokens each of which also occupy a unique
position in the text. Therefore, when one speaks of the indi-
vidual sentences composing the text, it is to be understood as
merely a convenient manner of speaking.

Sentences containing non-empty referential occurrences of
constituents, such as relative pronouns, tense forms, words
like "this", "now", and "Tuesday", and pronouns like "I" and
"you", are called token-bound sentences. They are called
token-bound sentences because the specific interpretation of
the correct objective referents of such sentenial constituents
- according to their operational definitions - is bound to a
specific sentence token. Token-bound sentences comprise by
far the vast majority of sentences whose tokens are used in
any kind of discourse. These are the kinds of sentences whose
tokens can undergo a meaning change depending upon changes in
discourse context. This individual non-recurring meaning is
called the cognitive meaning of a sentence; the cognitive
meaning of a sentence is thus a function of a purposefully
used utterance token of an isolated sentence already possessing
constant linguistic meaning plus its position relative to other
utterance tokens of the discourse. The sentence given above
is an example of a token-bound sentence since its cognitive

meaning, as well as its truth-value, can be discovered only
when it is seen as an individual constituent in a context of
actual discourse. In some kinds of sentences, the cognitive
meaning never varies from the linguistic meaning.

The value of distinguishing between the linguistic mean-
ing and the cognitive meaning of a sentence for the purpose
of semantic theory is very great. It allows one to speak
unambiguously of the "meaning" of a sentence when that constant
linguistic meaning itself is under semantic analysis, as, for
example, when it is being related as linguistically synonymous
with the meanings expressed by other sentences under conside-
ration, such as "He was said to have robbed the store". The
cognitive meaning, which so often depends upon the arrangement
of the sentence token in discourse, can be anaylyzed success-
fully only after a thorough semantic study of linearly ordered
discourse has been carried out, i.e., when much more has been
learned about how the individual relative pronouns, tense forms,
etc., function semantically as they operate relative to one
another when they appear in the broader context of purposefully
chosen segments of discouse. This linguistic study is a study
that has scarcely been looked at since most grammarians and
semanticists have confined their attention to the syntactic
structures and meanings of isolated sentences.

This restriction of attention to the isolated sentence
has led to needless controversies about what is the so-called
form of certain English sentences. For instance, with respect
to a disagreement about a correct tense form in the discussion
preceding the presentation of this paper, it was pointed out
that   "He eats"   is not the correct from of this English
declarative sentence since we "normally" would say "He is eating"
when the sentence occurs as a single utterance. It was quickly
pointed out by those participants more accustomed to dealing
with sentence types appearing in discourse that "He eats" is

a perfectly well-formed English sentence used on occasions
when "He is eating" would be incorrect. Here is an example
of the two uses: My father has been very sick of late. He
eats. But he doesn't know what he is eating. He talks.
But he doesn't know what he is saying.

The problem of defining what the characteristics of a
sentence of the language are and where its boundaries are
to be drawn has long been one of the most troublesome problems
in linguistics. A prominent school of thought in contemporary
linguistics claims that no definition of a sentence, operational
or general, can possibly be given preliminary to the explicit
construction of a grammar, specifically a transformational
generative grammar. The position taken is that the concepts of
sentence and sentence boundary must be taken as primitive
since the potential infinity of the well-formed sentences of
a language, of necessity, can be recognized only intuitively
by the human language user who must already have mastered such
a grammar. Apart from the epistemological question whether
a theoretical linguist must necessarily accept the particular
assumptions leading to this point of view, it is clear that
this view holds little interest for those of us attempting to
solve the practical problem before us. We have to be able to
describe physically observable criteria explicitly enough to
instruct a machine how to determine mechanically the bound-
aries of any given sentence so that it can carry out further
explicit instructions how to determine its linguistic meaning.
The machine of course has no inborn intuition; and even if a
transformational grammar of the type envisaged by this school
of thought were incorporated somehow into the operating
capabilities of a machine, it is not feasible to demand of the
machine to generate sentences, by applying the ordered deri-
vational rules of a formal transformational grammar to the
vocabulary of a language, until it finally generates success-
fully a sentence whose symbols match exactly those of the

expression under consideration in the text, thereby assigning
to the expression the formal structural description supposedly
sufficient for determining its so-called "semantic interpretation".
The number of potential sentences generatable before success is
reached is far too great to imagine carrying out such a mechanical
procedure for a single sentence, much less carrying it out thou-
sands of times for each sentence in a text. It is no wonder that
many of those who hold such generative grammars - restricted to
sentential construction rules alone - to be the only correct form
of grammar also hold that any attempt to solve the problem of
instructing a machine how to recognize the information expressed
through the ordered discourse of the input text is foredoomed
to failure. Nonetheless, even if we are not daunted by this
pessimistic point of view, it is clear that if we are to succeed
at all we must discover some useful and yet theoretically
satifactory definition for sentences which provides mutually
acceptable operational criteria for determining their observable
syntactic and semantic characteristics.

The operational definition of a sentence of a natural
language, as proposed by this theory of structural semantics,
is the following: A sentence will be that linguistic expression
that can be shown empirically to convey at least one abstr t
sentential meaning recognizable to a reliable sampling of the
native speakers of that language. Moreover, only those expres-
sions which can be shown empirically to have this sentential
semantic property are to be regarded as natural language
sentences, because it can be demonstrated to be a necessary
property of natural language sentences, i.e., if any sequence of
conventional symbols is to be capable of expressing a message at
all, it must express this underlying abstract sentential meaning.

Before an illustration of a typical abstract sentential
meaning recognizable to fluent speakers of English is given,
the following general remarks can be made. The abstract sen-
tential meaning is an observable property that belongs only to

the sentence itself as an internally interrelated complete
semantic entity; to put it in a different way, it 's a kind
of meaning with a message content that is not transmittable
through any one of the component parts of a sentence. It is
also a semantic property belonging to that kind of isolated
expression which we more or less all mutually agree belongs
only to those expressions linguists have habitually regarded
as complete and well-formed sentences. Hence the underlying
abstract sentential meaning accounts for the so-called inborn
intuitive recognition of a well-formed sentence, postulated by
N. Chomsky as explaining the tremendous overlap of agreement
among fluent language users as to which expressions constitute
a set of representative sentences of the language. It can be
proven that we must comprehend this meaning before we can
recognize the linguistic meaning of a sentence, yet this kind of
of sentential meaning has not hitherto been explicitly recognized
as a significant, universal, observable semantic phenomenon.
There is abundant evidence that the existence of the abstract
sentential meaning has been implicitly recognized by linguists,
even generative grammarians, because it can be shown that they
consistently make implicit use of the recognition of abstract
sentential meaning as a discovery procedure for establishing
the significant syntactic characteristics of a language, but
no systematic attempt has been made to specify its exact
characteristics. Logicians too have attested to its existence
when the various needs arose. A good example in the English
language is the hypothesis contrary-to-fact type of sentence,
which has undergone much discussion by contemporary logicians
because the obvious abstract sentential meaning it expresses
cannot be formulated within the specific linguistic techniques
provided by the purposefully restricted formalized languages of
deductive logic. Yet no logician has ever analyzed the essential
characteristics of this particular type of sentence whose

expressed abstract sentential meaning gave rise to its
own aptly decriptive name. Therefore, for the sole purpose
of illustrating the semantic phenomenon now under discussion,
let us look at the following sentence: If Germany had invaded
England, Germany would have won the war.

When a sentence of this type is declared to be true,
it is immediately understood as asserting certain significant
facts about its own subject matter: First, neither of the
two events specified in the sentence has happened in actuality
the two events mentioned having been specified through the
descriptive terms: Germany, invade, England, win, and war,
terms which function semantically to determine what is called
the referential context of the sentence. Second, the event
mentioned first is a sufficient condition of the event
mentioned second. How are these particular facts conveyed?
It is not possible for this kind of information to have been
conveyed through the referential context because every one of
the descriptive terms can be replaced by other members of the
same syntactic-semantic categories the original descriptive
terms belong to. Thus: If George had married Jane, he would
have bought the house, is another sentence which expresses the
very same abstract sentential meaning although its linguistic
meaning is utterly different.

Inspection of the two sentences shows equally well that
neither statement contains an explicit statement of the
expressed abstract sentential meaning. How do we know that
neither event occurred? One observes the interesting semantic
phenomenon that nowhere in either clause does there occur a
negative particle, such as not, explicitly denying the existence
of either event. Indeed, had there occurred a not in either
of the clauses, the sentence would have conveyed the abstract
sentential meaning that the event mentioned in the clause did
in fact occur!

Furthermore the abstract sentential meaning expressed
by the sentence as a whole does not depend upon a predeter-
mined "meaning" communicated by the particle if. As an
illustration, the identical if-clause of the sentence immedi-
ately above can appear in a third sentence: If George had
married Jane, he had divorced her too, which expresses the
very different abstract sentential meaning that George did
in fact marry Jane and did in fact divorce her. Moreover,
in this last sentence, no sufficiency condition is maintained
as relating the two events as causally connected; they are
truth-functionally related, i.e., the interpretation of if in
this sentential context is what traditional philologists have
termed the concessional use of if.

This last sentence illustrates very nicely why the
language user has to comprehend the whole of any sentence as
a complete semantic entity before he can determine its correct
linguistic meaning. Even more significant, the two examples
immediately above illustrate that before the formal linguist
can determine what are usually regarded as the purely syntactic
features of either of the two sentences, he has first to recog-
nize each different abstract sentential meaning underlying each
linguistic meaning respectively; as one instance of now the
linguist uses this information given through the whole sentence
as a discovery procedure to differentiate among identical
constituents when they function differently in different senten-
tial contexts: in the sentential context of the first sentence,
the constituents had married have to be interpreted as a
syntactic form of the verb marry in the subjunctive case; in
the sentential context of the second, the same constituents have
to be interpreted as a syntactic form of the verb marry in the
ordinary past perfect indicative case. Moreover, the constituent
would appearing in the sentential context of the first sentence
does not alone account for the interpretation of its abstract
sentential meaning as expressing an hypothesis-contrary-to-fact,

as suggested by Y. Bar-Hillel during the ensuing discussion.
This fact can be demonstrated empirically by the construction
of a sentence with an occurrence of <u>would</u> in the main clause,
a sentence which expresses yet another abstract sentential
meaning: <u>If George had to study, I would have my girl friends
to tea.</u>

Through what language devices, then, is the abstract
sentential meaning expressed? According to the proposed theory
of structural semantics, it is expressed through and only
through the ordered combination of all of what will be called
the <u>structural</u> <u>semantic</u> <u>properties</u> of the sentence. The
structural semantic properties are defined as those, including
all occurrences of non-descriptive terms appearing in their
original constituent order, which remain in the sentence when
each of the denotative morphemes (stems or words belonging to
abstract grammatical classes) has been abstracted out and
replaced by the syntactic-semantic category of which it is
a proper member.

As an illustration, the structural semantic properties
of the two hypothesis-contrary-to-fact sentences are exactly
specified by the formulation: A <u>If x had j-ed y, x would
have h-ed the z</u> (where A is a structural semantic symbol
introduced to represent <u>declarative</u> <u>mood</u>, expressed in English
by intonation and constituent order; x, y, and z are variables
ranging over nominals with different cases; and j and h are
variables ranging over verbals). This kind of linguistic
formulation is called a <u>sentence-abstract</u>. A sentence abstract
is a structural semantic formula that exactly specifies what
is called the <u>structural</u> <u>semantic</u> <u>context</u> of a sentence. The
structural semantic context of every sentence must be carefully
distinguished from its referential context which is supplied
only when the variables have been specified so that the sentence
has a linguistic meaning. Thus, the linguistic meaning of a
sentence is a function of its abstract sentential meaning plus
its referential context.

Sentence-abstracts can be likened to the symbolic
formulas of a logical language-system, exemplified by the
well-known formula: $(x)/\overline{f}(x)\supset g(x)/7$, to be read "for every
x, if x is f then x is g". However, sentence-abstracts
cannot be identified with such formulas if for no other
reason than logical formulas are composed of ideograph-like
symbols. The structural semantic contexts of natural language
sentences specified by sentence-abstracts exhibit not only more
complicated inter-related structures than do the symbolized
logical formulas, they syntactic and semantic characteristics
are also more exactly specified. The logical formulas, when
they are used by language users and hence their symbolized
forms have been exactly translated into the specific linguistic
techniques and expressive forms of a specific natural language,
form a proper subset of the natural language sentence-abstracts
of that language. From the viewpoint of a theoretical linguist,
the formalized symbolic languages are restricted language
systems, ingeniously isolated out from the larger context of
the natural language systems and refined for very special
purposes. Thus they are special-purpose languages whose
sentences are imbedded within those of a natural language; the
formalized languages therefore are not approximations, in any
significant theoretic sense of this term, to natural languages.
The concept of sentence-abstract is thus, in this sense, a
generalization of the concept of the linguistically-interpreted
logical formulas. It should also be pointed out that if a
language user - an all of us are language users - did not have
the mastery of his own natural language he would not be able
to use the special-purpose formalized languages successfully
since the structures of these languages have been so specified
that their rules hold only for very restricted kinds of
declarative sentences, which express correspondingly very
restricted abstract sentential meanings.

The observable characteristics of the structural semantic contexts of various types of very fundamental sentence abstracts have been learned by the language user during the corrective feedback process of his learning period, and the abstract sentential meaning each such context expresses is recognized quite unconsciously during his participation in discourse. Thus, just as much as laws of nature, the fundamental sentence-abstracts have to be discovered and tested for correctness on the basis of empirical observation. These sentence-abstracts thus are basic forms which can be purposefully expanded to express new and different abstract sentential meanings by the application of construction rules determining when certain of its parts can be replaced by more complicated sentential sub-expressions, such as when a descriptive phrase can take the place of a simple noun form. Hence no general definition of tne concept ot abstract sentential meaning can be given. The explicit formulation of the working definition is: The abstract sentential meaning must be cognitive information that is completely expressible through the structural semantic context of the sentence and must be indubitably recognizable and agreed upon by a reliable sampling of fluent fellow language users when so formulated.

The recognition of the abstract sentential meaning is the sine-qua-non-ical condition of understanding the linguistic meaning of a sentence and hence its congnitive meaning. It is of course not sufficient to achieve a full recognition of the linguistic sentential meaning because the referential context must also be comprehended. However, the denotative morphemes can function to help express an individual message only when they appear within the complete framework of the structural semantic context of the whole sentence. By themselves they cannot contribute to meaningful discourse because they do not express a recognizable message when they are seen in isolation; they are not yet what we call language.

There are two, and only two, main kinds of structural
semantic properties in every sentence-abstract. The one
kind are called <u>structural constants</u>; the other kind are
regarded as <u>purely formal syntactic properties</u>. Both kinds
function together as inseparable expressive properties within
a unified context, both equally essential in producing the
abstract sentential meaning. The structural constants however
are regarded as semantic in character and must be distinguished
from the purely formal syntactic characteristics for several
important reasons. Structural constants are morphemes like
<u>all</u>, <u>even</u>, <u>not</u>, <u>only</u>, <u>ever</u>, <u>any</u>, <u>every</u>, <u>would</u>, <u>can</u>, <u>the</u>, <u>a</u>,
etc., all tense forms, all connectives, all expressions of
the imperative, declarative, interrogative moods, and some
perhaps not yet identified. They are expected to number in
English approximately a hundred. They are similar in function
to the logical constants of a typical logical system in that
they function as operators. They are said to have an <u>opera-</u>
<u>tional meaning</u> in contradistinction to the <u>syntactic signifi-</u>
<u>cance</u> of the formal syntactic characteristics and the <u>descrip-</u>
<u>tive meaning</u> of descriptive terms.

The operationally testable distinction made between the
two kinds of structural semantic properties is based upon the
very different functions each kind performs in contributing to
the recognition of the abstract sentential meaning. It can
be shown that the operational meaning of the structural constant
always enters into the information or message content of the
expressed abstract sentential meaning and hence into the
linguistic sentential meaning where the denotative morphemes
play their part. Thus, it is obvious that it makes a difference
to the objective and verifiable information transmitted whether
<u>A man came</u> or whether <u>All of the men came</u>.

On the other hand, the syntactic characteristics of a
language never enter as an integral part of the messages expressed

through the sentences in which they appear. They are the
morphological properties that express to the language user
the correct formal organization of the differentiable symbols
composing the sentence; the language user needs to know the
form of this structure if he is to convey and comprehend
information successfully. The overall function of syntactic
structure is to inform the language user how to coordinate the
organized structural semantic form of the sentence itself -
whose morphological characteristics are very different
physically from the physical characteristics of the world
about us - correctly to whatever it is that is being talked
about in that extra-linguistic world.

The relation of the sentence, which is an internally
organized linguistic entity, to objective reality is not a
direct one of naming or denoting physical entities, such as
so-called objective facts or events; the intellectual process
of coordinating the messages expressed through the physical
linguistic entities that are sentence utterances to the
reality talked about is thus a much more complicated process
than the mere direct pointing to "something", be it "truth-
value", "intentional meaning", or a vaguely defined "propo-
sition". The body of the formal syntactic rules of the lan-
guage implicitly supplies the necessary information as to how
a correct coordination to objective reality is to be made in
that particular language, much as the legend of a particular
map explains explicitly how the physical features of the map
itself are to be interpreted as depicting faithfully an actual
geographical area, so that the information expressed through
the physical characteristics of the map can be used directly
by the human as useful for establishing future plans for
exploring the actual area.

When important syntactic rules are broken in the construc-
tion of the morphological shape of a given sentence, there is

no possibility whatsoever for the resulting sequence of symbols
to express a recognizable meaning. However, when specific rules
governing the correct occurrences of structural constants alone
are violated, a special kind of logical contradiction can ar
a logical contradiction quite different from the kind of cont
diction that results when two descriptive terms occurring in a
given sentence contradict one another, as in the case of a
"round square". Thus, when one says: After John drink any milk,
he went to school, the logical meaning that obtains from the
first directive of ordering two events in time as one before the
other - such as in this case where the event of John's going to
school is directly specified as necessarily occurring after the
event of John's drinking an amount of milk has been completed -
is incompatible with the second directive given by the use of
the structural constant any. This inconsistency occurs becau'
any operates to express the further directive not to set an
upper limit on the amount of milk that has been drunk whereas
the amount of milk drunk has necessarily had to have a limit -
even if the exact limit itself has not been specified - because
the event of drinking of the milk has been described as neces-
sarily preceding the event of going to school. Therefore, the
inacceptability of this s quence as a well-formed, message-
bearing sentence derives from the fact that some of its operators
give incompatible directives. This conflict of directives can-
not be tolerated as a proper application of the rules governing
the correct use of structural constants in constructing a
message-bearing sentence since the directives given by the
operator-like structural constants always have to be logically
         with one another within the structural semantic fra
       the sentence if a consistent abstract sentent'
      be expressed.
      It should be noted that there are no
reasons whatsoever for ruling this sequence out from we

sentences because one can construct a well-formed message-bearing sentence which is syntactically similar to the first: Before John drank any milk, he went to school. The fact that the directives given through the operational meanings of structural constants may be inconsistent within the structural semantic framework of a sentence is taken as another important reason why structural constants have to be distinguished carefully from purely syntactic features. Purely syntactic forms are never logically inconsistent in this way. Inconsistent uses of tense-forms in structural semantic contexts, such as in: After John went, I will go, also exemplify this type of structural semantic inconsistency, a fact which demonstrates that the rules governing tense-forms are also not purely syntactic in character.

An important kind of intra-linguistic sentential translation law can be formulated when it can be established empirically that two or more sentences whose referential contexts are identical but whose structural semantic contexts are morphologically unlike one another, express the same recognizable abstract sentential meaning. For example, the sentence given above: If George had married Jane, he would have bought the house expresses the same abstract sentential meaning - and hence the same linguistic meaning since the referential context is identical - as the sentence: George would have bought the house only he did not marry Jane. Note that the second clause after only occurring in the last sentence states explicitly the fact that George did not marry Jane by the use of the indicative form of the verb and the direct use of the negative particle not. These sentences are said to be structurally synonymous to one another. Structural synonymity is of necessity a sentential relation - a generalized structural semantic relation of which the logical relation of equivalence is a special case - since it holds only between the sentence abstracts or different sentence types regarded as whole semantic units where there

exists no one-to-one morphemic synonymity. It is a relation
that is transitive, reflexive, and symmetrical in the logical
definitions of these terms.

Structural synonymity laws are important for the theory
of interlingual mechanical translation in that they are use-
ful for establishing interlingual sentence-by-sentence
translation laws. The concept of the abstract sentential
meaning thus serves as one of the technical semantic links
which enable the interlinguistic translator to map structural
semantic contexts of fundamental sentence-abstracts belonging
to sentences of one natural language system onto the laws of
another natural language system even though their grammatical
systems differ in every respect so that no word-by-word transla-
tion could possibly be carried out successfully. Empirically
established sets of structurally synonymous sentence-abstracts
belonging to each language system can be coordinated as sen-
tentially structurally synonymous to one another if the dif-
fering structural semantic contexts belonging to each set from
each language all express the same abstract sentential meaning.

The preservation of the correct abstract sentential mean-
ing is the conditio sine qua non of correct translation. Thus,
establishing empirically which structural semantic contexts of
sentences express the same abstract sentential meaning in the
differing natural languages is a necessary first step if one
is to achieve accurate translation from one language system
into another.

In the context of connected discourse, the referential
context extends beyond the confines of a single sentence.
Ambiguities owing to difficulties of determining the correct
objective referent of a member of a given syntactic-semantic
category, such as a noun-class, because the descriptive terms
may alike refer to different objects, may be resolved some-
times by taking increasingly larger segments of the discourse

surrounding the isolated sentence so that an inspection of this larger segment of referential context may determine the correct objective referent. The discourse segment should be chosen as small as possible, extending it only when necessary.

On the other hand, if the structural semantic context of an isolated sentence is ambiguous in the sense of expressing more than one recognizable abstract sentential meaning, inspection of an increasingly larger segment of the discourse context may also resolve the ambiguity.

The justification for this procedure of taking first the smallest possible segment of discourse and extending it only when necessary, is that it is to be expected that the discourse environment closer to the ambiguous expressions has the greater influence on the cognitive meaning since otherwise too great a strain on the memory capacity of the human might arise. Of course the machine will be expected to store some of the information accumulated during the chronological progression through the text.

DISCUSSION

ULLMANN: This is more a terminological question that a substantive one, but I am just wondering whether the dichotomy which you suggest between structural and referential is really a dichotomy; whether there isn't a third possibility, and also whether, if there is one, we shouldn't preserve the structural for that one. I am thinking of the passage in Chomsky's last book, "The Aspects," where he says, in addition to the sort of referential or denotational definition, which is part of the Katz-Fodor dictionary, there are what he calls "field properties," conceptual spheres.

CHARNEY: Then he would have to define what a field property is.

ULLMANN: Well, yes.

CHARNEY: If I should say, then, one can set these up, this is an empirical problem. If you want to say this is a field property, I would be quite willing to say it.

ULLMANN: No, I don't mean that. I mean the sort of field that surrounds the operation, the circumferential analysis -- kinship, intellectual terms. There are quite a number of those, and current usage has now very wisely preserved the term "structural semantics" for that. So what I am a little bit afraid of is that there might be terminological confusion if we called the very meticulous study that you are advocating "structural semantics." I would rather call it something else; I don't know what. Mr. Weinreich spoke of "compilatory semantics." That's one possibility, or "sentence semantics," or whatever. I wouldn't say it was a dichotomy and I wouldn't use "structural semantics."

CHARNEY: I think in one way the dichotomy has to be set up, and that I can argue perhaps on a technical level later. That would be hard to go into.

So far as the name is concerned, I found out myself after I had adopted this term that the book by John Lyons had appeared, and it was called "Structural Semantics," and I had a lot of soul searching to see whether I should go on using the term.

ULLMANN: He didn't introduce the term, but he used it.

CHARNEY: That's right. He used it in a sense that is extremely different. I don't like also to introduce a lot of terminology. I like to have it simple and just use the old terminology. Perhaps one could call it, instead of "structural semantics," "logical semantics," because what happens is that actually we are in the realm of logic and the sentence abstract is a very broad general sense of a formula that you get in a logical system, except that the logic is much more restricted. They are not able to handle all of the sentence types that are of interest and of importance to a natural language.

But again, if I use the word "logic" then it can be confused with the formalized term, so I would welcome any suggestions.

LEHMANN: I would like to ask, in conjunction with Professor Ullmann's remarks, whether Miss Charney isn't actually doing a type of field theory of sentences in the same way that has been done of individual things like colors, and so forth.

CHARNEY: It would be different, because color and so on have a kind of relationship. This is a theory to explain how we understand the meaning of a sentence.

LEHMANN: You are setting up sets of sentences, you see, in somewhat the same way, in much the same way, as people have grouped "red", "green", "blue"and so forth together.

CHARNEY: You see, I have given you a small example. On the basis of the structural laws, and so on, one can, let's say, reconstruct the logical system underlying the natural system. This has not been done. We could even find some that are more fundamental than others. "If" is a very fundamental one; "unless" is not as fundamental. The logicians selected very powerful operator words, which I call structural constants, except that structural constants are a wider class while the others are special cases. I don't want to say one word against logic; we understood much more about the language having seen what they did and the expressiveness that they were able to have, and they did distinguish between operator words and descriptive terms.

GARVIN: I wanted to make a very trivial comment in regard to the terminology. If I remember correctly, Priest used to differentiate between structural meaning and referential meaning in just about exactly the same way what you do, and perhaps it might be useful to discard the term "structural semantics" for what you are doing in terms of what Professor Ullmann has said, that the term "structural" has a meaning which has hallowed tradition.

CHARNEY: I'll tell you why. There is a very good reason; now we have fused semantics with syntax, and it is structural semantics, whereas in syntax,there is such a thing as structural meaning, and that is purely syntactic. When we look at a form and recognize it as a well-formed string arising f   a formal grammar, this certainly has significance. It is   essed to us by its shape and by its form, so this is what   .all structural meaning.

Structural semantic meaning is different, and this is because also this structural meaning is, I think, a special sub-class of these.

WEINREICH: You have exemplified the difference between the structural and the referential elements in your illustrative sentence, but would you be able to give it a definition? If I wanted to study structural semantics in your sense, what segments of a sentence should I consider?

CHARNEY: I would never consider a segment of a sentence. A sentence is the smallest unit capable of conveying an abstract meaning.

WEINREICH: But you have selected some parts of the sentences for our consideration.

CHARNEY: No.

WEINREICH: You have substituted variables, or something.

CHARNEY: This is not substituting. I should have made it clear. If you were interested in referential semantics, then you would be interested in the relationships among these referential terms, like "glass" and "house."

WEINREICH: How do I know what is a referential term?

CHARNEY: This is something that is found by testing and observation. You simply remove it and you find it doesn't convey a meaning.

WEINREICH: You have produced a sentence abstract from a sentence. Could we go further and for "If" put down "W", and have "WXFY"?

CHARNEY: I see what you mean. I have used the terms them-
selves, rather than introducing a symbol for it. This I have
to introduce a symbol for, because in the English language
we have no part of the physical vocabulary that belongs to it,
and in the structural semantic context the order is very
important. The shape is extremely important as is every
item that goes with it. And there is no general definition
for an abstract sentential meaning, because it is something
that we understand or that we intuit. But you can lay down
the conditions when you have it. The conditions are four,
and they would be too technical for me to go into. But they
are defined.

IV.

SEMANTIC ALGORITHMS[*]

by

Margaret Masterman
Cambridge Language Research Unit
England

(Preliminary Discussion)

The purpose of the paper which I want here to present
is to make a suggestion for computing semantic paragraph
patterns.

I had though that just putting forward this suggestion
would involve putting forward a way of looking at language
so different from that of everyone else present, either from
the logical side or the linguistic side, that I would get
bogged down in peripheral controversy to the extent of never
getting to the point. I was going to start by saying,
"Put on my tomb: 'This is what she was trying for'." But it
is not so.

I don't know what has happened, but I don't disagree with
Yehoshua Bar-Hillel as much as I did.

And on the linguistic side I owe this whole colloquium
an apology and put forward the excuse that I was ill. I ought
to have mastered the work of Weinreich. (1) I am trying to.
But it is not just that simple a matter to master a complex
work in a discipline quite different from that which one
ordinarily follows.

I may misinterpret, but it seems to me that the kind of

suggestion I put forward in this paper could be construed
as a crude way of doing the kind of thing Weinreich has asked for.
Similarly, Elinor Charney: I think in a crude way we in C.L.R.U.
make a distinction analogous to the one you made, but I am
not quite sure I have understood your work correctly. I have
come here to get educated on these things.

I have some exhibits,* and there is not time to hand them
out. I mistook; I thought the conference would be smaller.
However, I will put them on the table--not under it--because,
after all, the title of this colloquium was not the philosophical
title, "Logic and Language," but the would-be scientific title,
"Computational Semantics," and therefore I think it is fair at
any rate to put computer output in a visible place.

But Yehoshua Bar-Hillel is actually very right when he
wants to question all the time what real use the computer can
be in this field. So don't be misled by the size of this
output. In all the devices used except one, which is the one
I want to talk about, the computer is used above all as a
clerical aid. One should be clear, I think, in doing semantics
work, whether one could have done it without a computer and,
if not, in just what way the computer was a scientific or
clerical device.

Phrasings

The hypothesis from which we start, and which there
is almost no time to defend, is that the semantic unit of
language is given by intonational and phonetic data and is
not perspicuous from written speech. This semantic unit we
call a phrasing. I will start, therefore, by defining a
phrasing: A phrasing is a piece of utterance consisting of
two stress-points, and whatever intonationally lies between
them or depends on them.(2) In other words, phonetically
speaking, a phrasing is a tone group. (3)(4)(5)

To illustrate the nature of phrasings I give, as example,
the beginning of the last paragraph, phrased up by hand in a
rough and ready manner.

*See Ap endices A-F.

/The <u>hypothesis</u> ( )/	**Key:**
/from/<u>which</u> we <u>start</u> /	/ / boundaries of phrasing
/and <u>which</u> there+is <u>almost</u>/	( ) silent beat
/<u>no+time</u> to <u>defend</u> /	+ intonational connection
/ <u>is</u> that+the <u>semantic</u>/	____ stressed word
/<u>unit</u> of <u>language</u> /	
/is <u>given</u> by <u>intonational</u>/	<u>Note</u>: Segments smaller than
/and <u>phonetic</u> <u>data</u>/	the word are not here
/and+is <u>not</u> <u>perspicuous</u>/	stressed.
/from <u>written</u> <u>speech</u>./	
/I+will <u>start</u>, <u>therefore</u>,	
by <u>defining</u> a <u>phrasing</u>./	

You will appreciate that the phonetics of intonational form
is a definite discipline and that it is not the subject of dis-
cussion here. I can sustain discussion on what we at C.L.R.U.
are doing to make precise the study of what these phrasings look
like in actual text; but I give warning that this study will
involve further massive and tight experimentation, which we at
C.L.R.U. are not equipped to do.

Three lines are being pursued.

1. The Gsell Tune Detector at Grenoble will give the data.(6)
The technological difficulty in recording phrasings is that of
making static recordings of pitch data; and the Tune Detector will
do this. But even if literally miles of output were to be
obtained from such a tune-detecting machine -- and we do need
literally miles of output from it to allow for variations between
speakers -- this output would be very little good without the
possibility of subsequently processing it. We are therefore
struggling in C.L.R.U. to find a way of making a computer
simplification of it, so that the program itself (a clerical aid
again, but nevertheless a good one) can process this output
mechanically and analyse it.

Then, secondly, a statistical survey is being made of the characteristics of phrasings in English and Canadian French; these phrasings have been antecedently marked in the text by hand. (7)

Thirdly, there is one "hard" criterion of the existence of phrasings which I can here and now show. We have been examining comparatively large masses of official text issued by the Canadian Governmer. This has the original English and the Canadian French translation published together in the same volume. By examination of actual material, we have been trying to see what it would be like for a machine to perform the transformation from the English to the French. Such an examination exposes whoever makes it to the full shock of discovering the absence of linkage between any initial text and some other text which purports to be a translation of it in some other language. The sentential breaks do not always correspond; it goes without saying that the syntactic forms do not correspond, since a Frenchman translating from English takes pleasure in not letting them correspond; the vocabulary of course does not correspond. What hen does correspond? What corresponds is that the translation goes phrasing by phrasing. /See Appendix A/.

Since the phrasing proves to be so important, therefore, as the semantic unit of translation, my second exhibit, SEMCO, (8) /See Appendix B/ is the first output of a semantic concordance of phrasings which, in design anyway, is a considerable improvement on the I.B.M. Key Word in Context. (9) The merging and sorting program for this concordance is not finished yet; but it can already be seen from the output that the phrasings of which it is composed can each be sorted in three semantically significant ways: i) by the main-stressed word; ii) by the secondarily-stressed word; and iii) by the total unstressed remainder of the phrasing, or pendant. We hope to make this concordance a trans-lation-aid by setting it up bi-lingually: that is, by setting up

a set of correspondences between phrasings in English and
phrasings in Canadian French, and then programming a reactive
typewriter, on which the human translator will type out whole
phrasings in English, to do what it can to retrieve some phrasings
in the French. If the English phrasing consists of a technical
term or a stereotyped piece of officialese or an idiom, there will
be a one-to-one match with the corresponding phrasing in Canadian
French. If not, we hope progressively to enrich the system so as
to enable it to retrieve French translations of semantically
cognate English phrasings, i.e. either of other phrasings which
have both the same words stressed, but with different pendants;
or with phrasings with one stressed word in common with the original
phrasing; or with phrasings with the same pendant; or with
phrasings synonymous with the original phrasing in some defined
sense.

Thus, supposing that

/the <u>Queen's</u> <u>Government</u>/
/the <u>Canadian</u> <u>Government</u>/
/in <u>Canada</u> (   )   /

were all with their translations in the concordance, but

/Her <u>Majesty's</u> <u>Government</u>/

for some reason was not, the concordance would retrieve the
first two of these, in order of closeness, with their French
translations, on the ground that they had one common stressed
word with the original (namely, "<u>Government</u>") and that <u>Queen's</u>
is here synonymous with <u>Her Majesty's</u>.

Similarly, suppose the second phrasing, in the same text,
was

/is+of+the <u>considered</u> <u>opinion</u>/

the concordance might retrieve (e.g.) and in the following order:

/is+of+the <u>opinion</u> (   ) /
/has+given <u>serious</u> <u>consideration</u>/
/has <u>formed</u> the <u>opinion</u>/
/we <u>think</u>   (   ) /

In this case, the first of each of the two sets of retrieved phrasings, i.e., /the <u>Queen's</u> <u>Government</u> / is+of+the <u>opinion</u> (   )/ would indeed be a pretty good paraphrase of the original /Her <u>Majesty's</u> <u>Government</u> / is+of+the <u>considered</u> <u>opinion</u>/. But notice also that even in the worse case, obtained by taking the bottom phrasing of each of the two sets of phrasing retrieved by the concordance, some inkling would be retained in context of the brute sense of the original by saying

/In <u>Canada</u> (   ) / we <u>think</u> (   )/

All this is in the future, and we want to test it out in a pilot-scheme; in particular, we want to watch the concordance for size. What is already true is that we have made comparative analyses of quite a quantitiy of English and Canadian French text, including a text of 375 continuous phrasings, and there are only very few counter-examples to the hypothesis that you can go through, as in Appendix A, from parsing to parsing.

There is another point. A program is being written by John Dobson for marking phrasing boundaries from written text, using syntactic information. Some output from a dry run of the algorithm will be found in Appendix C. But, in fact, the phrasings do not always go with the syntax, though they usually do. See, for example, such English phrasings as

/A <u>man</u> who+is <u>said</u>/

/<u>Although</u> there+has <u>been</u>/

We have here two separable sub-systems operating within the total system of language: an intonational phrasing-system determining the semantic units of the message, and a grammatico-syntactic system, determining the grammatico-syntactic groupings of the utterance. They usually draw boundaries at the same places, but not always.

We can, of course, stress any segment of speech up to quite a long string of syllables. In that case the pace of speaking accelerates, though the rhythm not much. Here, as I have already

said, when any syllable has been stressed, I have underlined
the whole word; and I have used + signs to connect contiguous
stressed or unstressed words. I have also used empty brackets,
(   ), to denote silent beats or pauses.

I will not here discuss the notorious difficulty created
by the fact that different speakers stress the same passage
differently, except to say that in our so far limited experience,
the longer the text, the more unequivocally determined the
stress pattern.

Quatrains

The second semantic assumption which we make at C.L.R.U. is
that phrasings tend to couple up in pairs, and the pairs in turn
to couple up in fours.

Thus, taking again the last paragraph which I have written
and phrasing it up by hand in a rough and ready manner, we get

   /The second semantic+assumption/
   /which+we make at C.L.R.U./
1 /is+that phrasings tend/
2 /to couple+up in+pairs ,/
3 /and+the pairs in+turn /
4 /to couple+up in+fours./

or, said more quickly,

1. /The second semantic+assumption/
2. /which+we make at C.L.R.U./
3. /is+that phrasings tend+to couple+up+in pairs/
4. /and+the pairs, in+turn, to couple+up+in+fours./

These pairs of pairs of parsings, however obtained, we call
quatrains.

It is clear from the above example that this second assump-
tion is normative. In the case of a short piece of utterance,
in particular, one can always so arrange it that the phrasings
fall in fours, and one can, alternatively, so arrange it that

the phrasings fall irregularly. Moreover, this second hypothesis
is elastic, in that, to make it work, you have to allow for silent
beats. And though there is a consensus of opinion that these
genuinely exist, (10) there must obviously be independent criteria
of their existence and location for them to be usable in defence
of the quatrain-hypothesis; for otherwise, by just inserting up
to four silent beats wherever needed to complete a quatrain,
any piece of prose whatever could be analysed into quatrains.

I should prefer, therefore, to call the assumption that
there are quatrains a device, rather than a hypothesis. But it
is an extremely useful device, for by using it we can (and do)
provisionally define a standard paragraph as a sequence of four
quatrains, i.e., as a Quatrain. We can then suggest that into-
nationally speaking, the constituent quatrains of a Quatrain
(call them quats) may themselves be intonationally inter-related
by higher order phrasings, with higher order stresses, these
higher order stresses being spread over longer lengths of text,
thus producing a hierarchical intonational picture of a standard
paragraph, as illustrated on the next page.

STYLISED QUATRAIN-SCHEME OF A STANDARD PARAGRAPH

In this schema at each level, L, the main stressed segment of the first phrasing on the level below L, i.e., of the level L-1, becomes the secondary stress of the new phrasing in L, while the whole second phrasing of L-1 becomes the new main stress in L.

MEGA-QUAT, [or, SECOND-LEVEL QUATRAIN]

MESO-QUATS, [or, FIRST-LEVEL QUATRAINS]

QUATS, [or, ZERO-LEVEL QUATRAINS]

main stresses (1)

main stresses (1)

main stresses (1)

(1) (2) (3) (4) (5) (6) (7) (8)
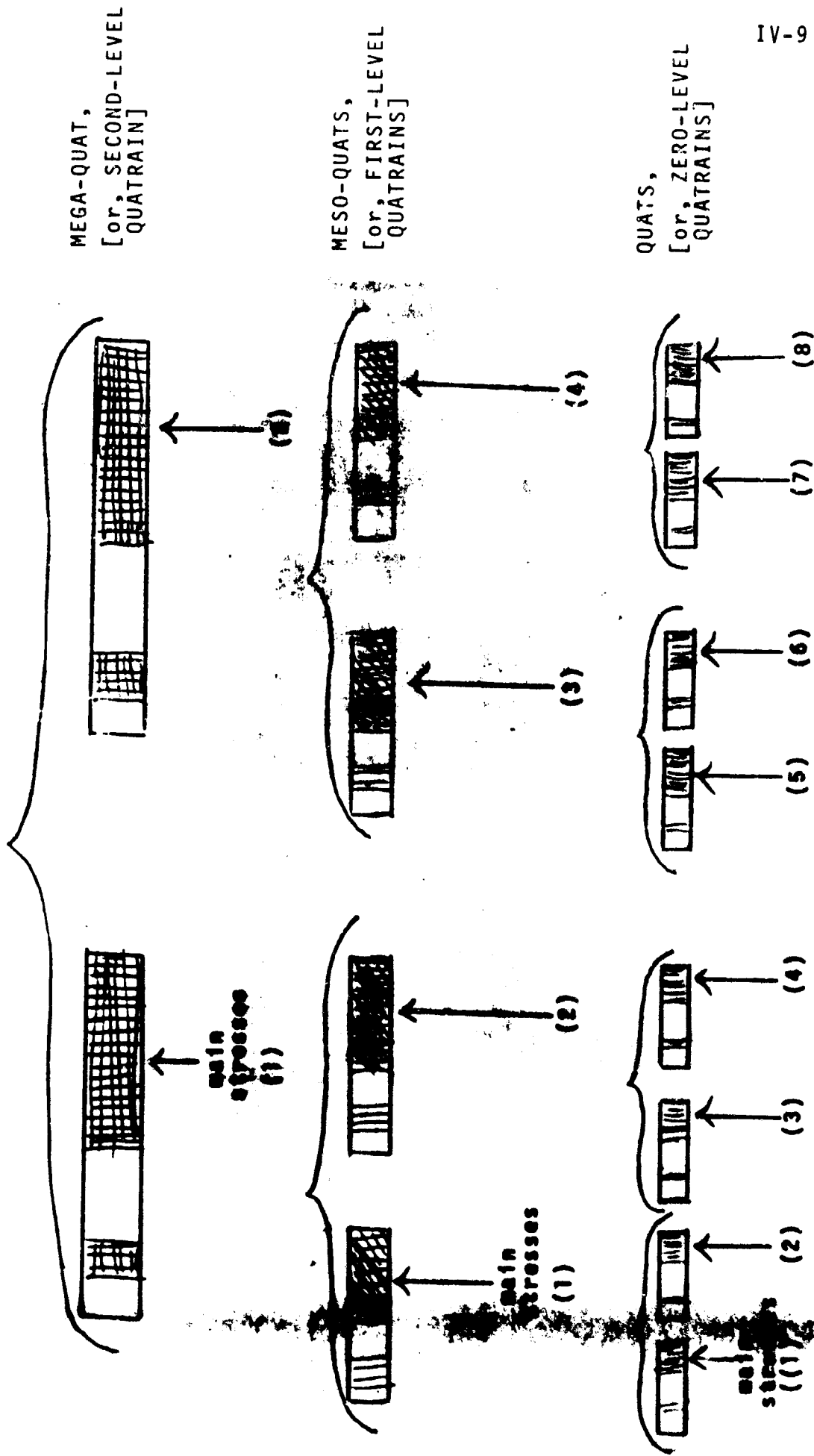
Of course this standard schema is a drastic and normative simplification of everything which intonationally happens in a real paragraph; it ignores all kinds of transpositions, aberrations and variants. Similarly, though more crudely, the hypothesis that a standard paragraph is a sequence of four quatrains itself tailors-to-shape any paragraph which is, in fact, not a sequence of four quatrains. But it is much easier in all study of language to analyse transpostions, aberrations and variants of anything if you have some initial schema or idea, simple enough to be easily grasped and retained by the mind, of what it is that they are transpositions, aberrations, and variants of.

This schema-notion also, you appreciate, like the phrasing hypothesis, constitutes the kind of provisional assumption that needs massive and precise experimentation. It ought to be possible, for instance, quasi-musically to estimate the accentuation or diminution of stressing which occurs in any segment of intonationally fully-contoured text according to whether the segment in question is or is not included within the boundaries of a higher-order stress. For instance, in the last paragraph which I have written immediately above (i.e., the paragraph which began "Of course this standard schema..."), my rough guess is that in the last sentence the secondary mega-stress of the final mega-phrasing is initial+schema+or+idea, while the main overall mega-stress of the same mega-phrasing, and therefore the intonational climax of the whole paragraph, is what+it+is+that+they+are+ transpositions+aberrations+and+variants+of; for note the tremendous emphasis, which I had to indicate by underlining even when writing down the original paragraph, of the final, usually totally un-stressed syllable, "of."

However, meso-stressing and mega-stressing are far away in the future. What I promised the organisers of this conference to bring along and try to explain were some exhibits of some C.L.R.U. semantic algorithms which had been used in the past. And I have

here some exhibits /See Appendix D7 which show the analytic use
we have made of the basic empirical fact on which the quatrain-
finding device rests, namely, that there is a sort of two-beat
rhythm (‖) which goes through discursive prose, especially
through the sort of discursive prose which occurs (e.g.) in the
London Times and in official documents:

1.	/A	man	who+is	said/
2.	/to+have	walked	through+the	ranks/
3.	/of÷the	Queen's		Guards/
4.	/	marching	through+the	Mall /
5.	/	taking		pictures /
6.	/with+a	ciné		camera,/
7.	/was	fined		£10/
8.	/at	Bow+Street		Magistrates'-Court/
9.	/	yesterday		( ) /
10.	/for	insulting		behaviour./ (12)

And in the 17th and 18th centuries, when prose was prose,
as it were, and a great deal of written text was composed to be
read aloud, the existence of this two-beat rhythm was deliberately
exploited. Here is the beginning of the philosopher Locke's
preface to his Inquiry Concerning Human Understanding:

1.   / I have put in thy hands /
2.   / what+hath been the diversion /
3.   / of   some of+my idle /
4.   / and   heavy   hours./

5.   / If it+has been /
6.   / the good+luck to   prove+so/
7.   / of any   of   thine,/
8.   / ( )   ( )

```
 9.  / and          then     hast+but  half/
 0.  / so+much      pleasure     in   reading/
11.  / as           I+had        in   writing+it,/
12.  /              (    )            (    )    /

13.  /              thou     wilt as little /
14.  /              think             thy money/
15.  /       as     I+do              my+pains/
16.  /              ill               bestowed./ (13)
```

Templates

    If the intonation of a paragraph is the study of its tune,
the semantics of it is the study of its pattern, because
the study of the kind of semantic pattern which occurs in a
standard paragraph has some analogy with the kind of pattern which
is mechanically searched for in pattern-recognition searches. I
have twice said (14) that in studying semantics one feels as though
one is idenifying a visual component in language rather than an
auditory component in language. This I should not have said unless
I was prepared to make it good, since such an analogy, being
as it is between two finite algorithms, must be by its nature
precisely determinable. I therefore do not wish to go any further
into this matter here, since it needs a special publication on
its own, which I hope in due course to provide.

    The reason that I was tempted to bring up this analogy at all
is that its existence (if it does exist) emphasises the main point
which I here want to make, namely, that formal logic as we at
present have it is not and cannot be directly relevant to the con-
textually-based study of semantic pattern. Logic is the study
of relation, not of pattern; and, in particular, it is the study
of derivability. By assimilating the kind of semantic pattern
which we in C.L.R.U. want to make a machine find with the kind
of visual pattern which research workers in the field of pattern-

recognition also want to make a machine find, I hoped that by
establishing a new analogy, based on visual pattern, I could
obliterate the thought of the false analogy between an applied
logical formula and a piece of natural language. But I see now
that I have been premature.

In order to get semantic patterns on to a machine, we have
created in C.L.R.U. a unit of semantic pattern called a template.
The word template, applied to natural language, has already quite
a history, having been used twenty years ago by Bromwich and
more lately by Miller. In the sense in which I am here going to
use it, it was a development of my earlier notion of a Semantic
Shell (15), simplified, streamlined, and further developed in
C.L.R.U. by Yorick Wilks. (16)

It will be recalled that a phrasing was earlier defined as
a piece of utterance consisting of two stress-points and whatever
intonationally lies between them or depends on them. Thus a
phrasing consisted, by definition, of three units, a main stress,
a subsidiary stress, and an unstressed part, or pendant.

I will try to make clear what I mean by the notion of a
double abstraction. The notion of a pendant is itself already
an abstraction from the linguistic facts because it creates one
unit out of one or more unstressed segments of text, which may
occur in the phrasing between the two stress points but may also
occur before or after them (also, of course, the phrasing may
contain no unstressed segment of text). Carrying this notion of
the form of a phrasing consisting of three units further, we
create three positions: an imaginary piece of metal with three
holes or template, the two end holes standing each for a stress
point, and the hole in the middle for the pendant, thus:

GENERAL TEMPLATE-FORM (STAGE 1)

0	0	0
1st stress point	pendant	2nd stress point
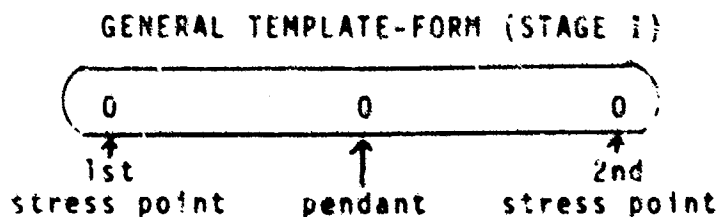
These units we fill with interlingual elements which, philosophically speaking, can be regarded as Aristotelian terms --indeed, though formally speaking they are not terms; they are in use the only genuine Aristotelian terms there have probably ever been. For an Aristotelian term has a) to be a "universal" (e.g., a general term like "pleasure", or "man"); b) to be such that two terms can be linked with a copula in between them; c) to be such that they can occur, without change of meaning, either as subjects or as predicates. As is notorious, this last is the difficult condition for an actual word in use in language to fulfil, for if we say, e.g.

"Greek generals are handsome."

using "handsome" here as a predicate, we have to continue

"Handsomeness is a characteristic of all the best men."

(or some such thing) if we are to use the term "handsome" also as a subject; i.e., we not only have to change its form, but also give it a far more abstract meaning than it had as a predicate. To use a semantic sign as a genuine Aristotelian term requires a quite new way of thinking. We achieve this be creating a finite set (~. 50) of English monosyllables of high generality (e.g., MAN, HAVE, WORLD, IN, WHEN, DO, etc.), and, divesting them by _fiat_ of their original parts of speech, ordain that they may be combined with two and only two connectives:

a) a colon (:) indicative of "subjectness"

b) a slash (/) indicative of "predicateness."

By using these two connectives we then recreate English "parts of speech" as follows:

Noun	a :
Adjective	a :
Verb	a /
Preposition	a /
Adverb	a / (17)

Finally, we rule that at least two terms shall be required to make a well-formed formula (the two terms having one connective between them), and say that any two-term formula ab in which the a and the b are separated by a colon (i.e., a:b) shall be commutative, whereas any formula ab in which the two terms are separated by a slash shall be non-commutative (i.e., a/b). Finally, a bracketing rule has to be made (not, I think, thought of by Aristotle), allowing any two-term formula itself to be a term. This set of rules for C.L.R.U. interlinguas has been given in various work papers and publications. (18)

Using this term-system, we fill in the holes in our template-form as follows:

$$(\underline{a}: \qquad (\underline{b}/ \qquad \underline{c})):$$

Since these brackets are invariant we may omit them giving

a:	b/	c:

e.g.

MAN:	CAN/	DO:

However, if it be remembered that a template is meant to be a coding for a phrasing, it is clear that we have now made a second type of abstraction from the linguistic facts. For we have not merely made a positional abstraction from them, representing the primary and secondary stress points, and the pendant of any phrasing. We have also, by inserting general terms into the three positions, made a semantico-syntactic abstraction from them; for a whole class of phrasings will, clearly, be representable by a single triad of terms

To separate the members of this class, we complicate our template by inserting into it three variables, α, β, γ, as under:

GENERAL TEMPLATE-FORM (STAGE II)

α a:	β b/	γ c:

These variables can be filled as values by further specifi-
cations, made by using the rules above, composed of terms; the
object of the specification being to specify the semantic content
of an actual phrasing sufficiently to distinguish it from all
other phrasings coded under the system which have the same
general template-form: (e.g.)

GENERAL TEMPLATE-FORM

α MAN: β DO/   TO

ACTUAL CODED PHRASING

(SELF:MAN)	(WILL/DO)	(CHANGE/WHERE)/TO)

PHRASING

/ I   will   come/

Sometimes, however well chosen the original set of terms,
thesaurus-heads or other descriptors are, in addition, necessary
to distinguish two phrasings from one another. (e.g.)

(ONE:Male MAN)   /(WILL/DO)/(CHANGE/WHERE)/TO)

/He will come/

(ONE:Female MAN):/(WILL/DO)/(CHANGE/WHERE)/TO)
/She will come/

It will be evident that, with so sparse a coding system, only a limited number of the shorter phrasings of natural language can be coded. For instance, I remember a long discussion in C.L.R.U. about how to code the phrasing

/that+it+was+the <u>Annual Fair</u>/

from the text "... then I found <u>that it was the Annual Fair</u>, which was always held at Midsummer...."

It is obvious that into this phrasing the information-content of two or more smaller phrasings taken from some such set as the following have been compressed, e.g.

/the <u>Annual Fair</u>/

/that it+was (   )/

/it+was <u>the</u> Fair (   )/

/(   ) it+was+the <u>Fair</u>/

/it <u>was</u> the <u>Fair</u>/

It is clear that it would not be out of the question to mechanize the process of cutting up one long phrasing into two small ones; but I do not want to go further into this here.

For what this query does is to bring up the far more funda-mental question, "What is this whole semantic coding technique for?" "What is it worth?" "And what is it going to be used for?" And it is this deeper and more philosophic question which I now want to discuss.

The <u>Semantic Middle Term</u>: <u>Pairing the Templates</u>

As I see it, in contemporary linguistics, there are two trends. The first is connected in my mind, rightly or wrongly, with such names as W.S. Allen, M.A.K. Halliday, John Lyons, R.M.W. Dixon, and of course, above all, J.R. Firth; and I there-fore think of it as "the British School of Linguistics," though it is almost certainly, in fact, a world-wide trend. The members of this school take raw untampered-with utterance and then try to segment it, analyse it, and account for it, using machines

as clerical aids but taking the text as given; they do not try
to add anything to it, excise anything from it, or otherwise
explain it away. They try, moreover, to name the categories
which they find from the operation of finding them, instead
of appropriating to new linguistic situations the well-known
hackneyed categories of Graeco-Latin grammar. The rationale
of doing this kind of work is brilliantly expounded in W.S.
Allen's Linguistic Study of Languages (19); and a major theoretic
work has recently been published from within this general trend,
namely R.M.W. Dixon's What is Language? A New Approach to
Linguistic Description. (20)

I will confess that it is with this school and not with
the M.I.T. school that my linguistic sympathies primarily lie;
for it seems to me that the whole point of doing scientific
linguistics -- the whole battle which it has taken the scientific
linguists thirty years to win -- is that the practitioners of
this technique engage themselves to open their eyes to look at
the utterances of the languages of the world as they really are;
instead of forcing them all (as in the older philology) into a
Latin-derived straightjacket; or seeing them (à la Chomsky) through
the distorting glass of an Americanized norm.

It is no accident, of course, that Allen and Halliday should
have formed my conception of linguistics, for W.S. Allen is
Professor at my own university, while M.A.K. Halliday, besides
being one of the group who originally founded C.L.R.U., also
put us on the original thesaurus idea, on which all our more
recent semantics work has directly or indirectly been founded. (21)
Also the view of language taken by the phonetic analysts, and in
particular by P. Guberina (22), much more nearly coincides with
that of "the British School of Linguists" than with that of the
present M.I.T. school.

But now we come to a difficulty; to another form of the same
difficulty which probably led Chomsky and his school, and probably

Fodor and Katz also, to make their drastic abstractions from the facts of language. If the distributional method of linguistics, unaided, is the only tool which is to be used to analyse and understand natural language as it really is, such language will remain forever unanalysed and non-understood; that is, it will remain ineffable. For even with a whole row of the largest imaginable computers to help, <u>all</u> the potential distributional potentialities of a <u>whole</u> national language cannot possibly be found in any finite time; (23) and it is part of the scientific linguists' contention that nothing less than the finding of the whole is any theoretic good. (24) Unless, therefore, some new technique can be developed, unless some fairly drastic abstraction can be made from the genuine linguistic facts so that a system can be created which a machine can handle and which has some precisely definable analytic scientific power, all the analytic linguists of the world will turn from truly linguistic linguistics back to Chomsky, Fodor and Katz (and now Weinreich), and they will be right.

Here I think I should do something to make clearer what the nature of my criticism of the Chomsky school is and what it is not. My quarrel with them is not at all that they abstract from the facts. How could it be? For I myself am proposing in this paper a far more drastic abstraction from the facts. It is that they are abstracting from the wrong facts because they are substracting from the syntactic facts, i.e., from that very superficial and highly redundant part of language which children aphasics, people in a hurry, and colloquial speakers always, quite rightly, drop. On the same level Chomsky wants to generate exactly the "sentences" of English; and yet, to do so, he creates a grossly artificial unit of a "sentence"; i.e., founded on nothing less than that old logical body, the p and q of the predicate calculus. (25)

Similarly, Fodor, Katz (and Weinreich), when doing semantics, talk about "contexts" and "features" and "entries in dictionaries";

but their dictionaries are always imaginary idealised diction-
aries, and their examples are always artificially contrived
examples, and their problems about determining context always
unreal problems. (26) So, for me, in spite of its clean
precision and its analytic elegance I think this approach
combines the wrong marriage of the concrete and the abstract.
That this is so is now beginning to be operationally shown, in
my view, in the appalling potential complexity which is about
to be generated by keeping all the transformations in the
calculus meaning-preserving, when the whole point of having
grammatico-syntactic sustitutions in a language at all is
that precisely they aren't meaning-preserving. And now that
the elephant of an encyclopedic semantics is about to be
hoisted on top of the tortoise of the already existent syntactic
Chomsky universe, it seems to me that the whole hybrid structure
is shortly about to topple with a considerable crash of its
own weight. And this is a pity indeed; for the complications
which have gathered obscure the whole very great potential
usefulness of the original, simple, and above all elegant,
analytic idea.

In contrast with this elegance, see the crudeness but
also the depth of what I now propose. I don't have sentences
at all: I have phrasings. And, granted also that in my first
model I can only have small phrasings (see above) and that I
can't yet distinguish differences of stress-and-tune within
them (see above) and that all my phrasings have to combine
in pairs; i.e., I can't yet accommodate triplets (see above);
and that the pairs of parsings have to be handled by a quatrain-
finding device (see above) which is itself highly artificial
and stylised (see also above), it yet remains true that, even
in my first semantic model, I can deal with stretches of
language like Trim's classic example:

$$/\sqrt{\underline{H-m}} \quad (\quad)/$$
$$/\diagup\underline{H}\ m \quad (\quad)/$$
$$/\diagdown\underline{Hm} \quad (\quad)/$$

let alone   /<u>Colourless</u> green+<u>ideas</u>/
                    /<u>sleep</u> <u>furiously</u>/

which Chomsky can't.

Secondly, I analyse these phrasings, even in my first model (see above and below) by a coding-device, which is philosophically derived not from the logic of predicates but from the logic of terms. This means that with fifty categorically-changeable operators, two connectives, and a bracketing-rule, I can create a pidgin-language, <u>the full structure of which can really be mechanically determined by the strict use of the scientific-linguistic methods of complementary distribution</u>; that's the cardinal point. (See Appendix E.) Maybe the first such structure which I propose is a wrong one: nevertheless, I alone propose some such structure.

Thirdly, even in my first model I make provision for the cardinal semantic-linguistic feature of anaphora (27), or synonym-recapitulation. Granted that syntactic interconnection in this model withers to a vestigial shred of itself, the far more cardinal rhythmically based phenomena of reiteration, recapitulation, and parallelism are centrally provided for. Likewise, with this coding the machine can write poetry and therefore handle metaphor; (28) though actual output from this has not been shown yet.

When it be considered, therefore, what, semantically, the C.L.R.U. semantic paragraph-model can do -- as opposed to what, grammatico-syntactically, it can't do -- a very different and much more sophisticated view of its potentialities becomes possible. This model is crude, yet: but its "deep-structure" unlike Chomsky's deep structures to date, is really deep.

And now it is necessary to show what that deep structure
is.

A preliminary remark: In judging it, it is necessary to
remember its technological provenance. It is just here, i.e.,
in the guidance given by technologies towards determining this
structure, as it seems to me, that the severe discipline
imposed on C.L.R.U. by sustained research in the technological
fields of Machine Translation, Documentation Retrieval,
Information Retrieval, and Mechanical Abstracting has stood us in
good stead. For the prosecution of these goals lends a hard edge
to thinking and an early cut-off to the generation of complexity
in programming, which purely academic studies of language do
not have. It was this technological pressure which led us to
Shillan's practical Spoken English (29) and the discovery of the
phrasing; to the semantic utilisation of the two-beat prose
rhythm, and to the quatrain-finding device and to the notion
that there might be comparatively simple overall intonational
contours to the paragraph.

And it is this same technological pressure which has
predecided for us what use we will make of all these stylised
and streamlined phonetico-semantic units. We code them up
into a crude but determinate "language," and then, by giving
this language vertebrae, as it were, i.e., templates, we
construct (or misconstruct) a paragraph's semantic backbone;
or alternatively, other parts of a text's semantic skeleton.

This is done by using the device of the "middle term."
The "middle term" derives in idea -- though not in use -- from
the syllogism as originally conceived by Aristotle (30), any
syllogism being here considered not as an inference structure
but as a text. Thus a syllogism, linguistically interpreted,
consists of three phrasings, which, between them, contain only
three terms; and the differing forms of the syllogism are dis-
tinguished from one another by reference to the action of the
middle term.

Here, analogously, we make the machine make a unit
consisting of two coded templates, the connection consisting
of the recapitulation of one of their constituent terms.

Thus if I code

/The	girl	was+in+a	house/
/and the	house	was+in+a	wood/
/and the	wood	was+full+of	trees/
/and the	trees	were+covered+with	leaves/

<div align="right">etc.</div>

I get templates of the form

α MAN:	β IN/	γ PART:
α PART:	β IN/	δ WHERE:
α WHERE:	β HAVE/	γ PLANT:
α PLANT:	β HAVE/	δ POINT:

and the recapitulation-pattern is as follows:
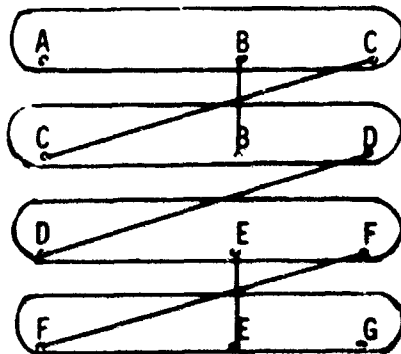
If, then, we further simplify by matching only "stressed" terms,
i.e., if we ignore as skeletally adventitious the recapitulation
of the two pendants in the middle positions, we are left with
what I believe to be one of the basic anaphora-patterns of all
language, i.e.
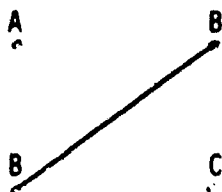
which, in the case of the syllogism, introduces the transi-
tivity-rule-carrying syllogism

"If <u>A</u> is <u>B</u>

and   <u>B</u> is <u>C</u>..."

It must be evident that, in terms of our model (and allowing
coded pendants as well as coded stressed segments to match)
we can have nine basic pairing-patterns.

Likewise, it will be evident that combinations of these
can be permitted (e.g., see above); and that, the set of 50
elements of the system being strictly finite, the strict matching
algorithm <u>A</u> <u>matches</u> <u>with</u> <u>A</u> can be relaxed to allow A to match with
some subset of other elements or with any other element. (31)

If only elements in the first and third positions are
allowed to match, we get four basic patterns, corresponding
indeed to the four categorical forms.
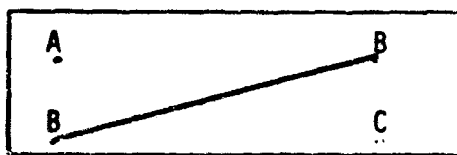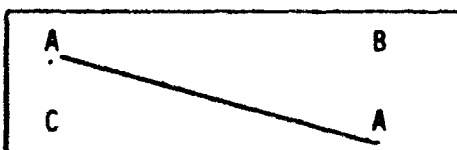
For this model -- and allowing for the f. ʾ that what has to
match are not, as above, the actual words of the phrasing but

the terms in the coded templates -- these are the four basic
semantic patterns of language.

The Philosophical Notion of the Semantic Square

It must be evident, from even cursory examination of the
above, that a great deal of meta-fun can be had, by inserting
a list of permitted pattern transformations into this model
to produce approximations to various brute syntactic forms;
or to account for ellipsis (which is only the same thing,
after all, as complete unstressedness); or, better still, to
make the machine infer "logical" interconnections between
various specifiable particular pairs of templates. This meta-
fun we in C.L.R.U. do not as yet propose to allow ourselves to
have. This is partly because having once broken right through in
our thinking, to a conception of phonetic-semantic pattern which
is independent of, because prior to, that of syntactic pattern,
we do not want prematurely to reimprison ourselves within the
patterns of syntax. It is also because we conceive our first
duty to be to try to put the machine in a position to proceed
from paired-phrasing-patterns to the overall semantic pattern
of a paragraph: i.e., not to find out what logically follows
from what, but, far more primitively, what can follow what.
To do this, we postulate a basic semantic pattern in language,
namely, Guberina's pattern of the "semantic square" (32) or
"carré semantique." This also derives from an "Aristotelian"
device; but I have caused a great deal of obfuscation and
confusion by stating, without further explanation and as
though the fact were obvious, that it derives from Aristotle's
Square-of-Opposition. (33) Psychologically, it does, and I
have no doubt in my own mind that in Guberina's case it did.
But to see how it did it is necessary to keep a basic hold on
three truths: Firstly, that the "Square of Opposition" forms
no part of syllogistic logic. Secondly, that it must be

reinterpreted for this purpose as being a logico-linguistic schema, giving a pattern of semantic contrast between four pairs of four terms. Thirdly, that it must then be generalised so that it can be restated as a semantic hypothesis, as giving the basic overall pattern of semantic contrast within a primary standard paragraph.

Thus the original Square of Opposition is a schema giving the valid forms of immediate inference between the four categorical forms:
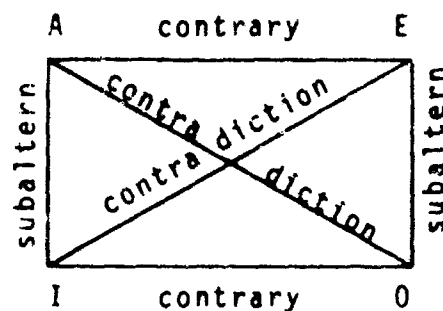
where A is: All As are Bs.

E is: No As are Bs.

I is: Some As are Bs.

O is: Some As are not Bs.

As is well known, when interpreted in terms of the logic of classes, or in terms of a logic of predicables, this schema runs into difficulties.

Interpret it now linguistically, i.e., in terms of the four following actual phrasings:

A   is: /All+As   are   Bs/

E   is: /No+As   are   Bs/

I   is: /Some+As are   Bs/

O   is: /Some+As are not Bs/

Now imagine other words in the stressed positions but keeping the semantic stress-pattern, so that realistic actual colloquial conversation results:

1st Speaker (a Scot)		/All+Irish are crooks./
2nd Speaker (an Irishman)		/No+Irish are crooks./
1st Speaker	/I don't care what you say./	/Some Irish are+crooks./
2nd Speaker	/And I repeat./	/And some Irish are utterly+ non-crooks. (i.e., the most honest characters alive.)/

Continue now the conversation in a realistic manner:

1st Speaker	/It comes to this./	/They're either+angels or devils / /Irishmen go+to extremes./
2nd Speaker	/Exactly./	/Some may+be utter+black+hearted+ fiends/ /but others are absolutely angels+ of light./

What on earth have we here? And in particular, what have we here if we reimagine this as a general standard of paragraph-schema, i.e., if we abstract from it by dropping the particular linguistic segments '"All," "No," "Not," "Some".' (For I am talking about the uses of these English words, not about logical quantifiers.)

What we have is a pattern of diminishing semantic contrast, which is accentuated by the necessity of constantly repeating all the terms (or rather if, by using the model, the phrasings were replaced by coded templates, the terms would repeat).

This pattern can be schematized as follows:
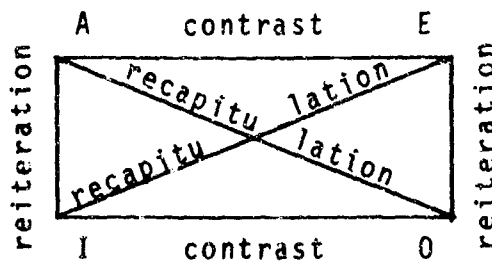
If we restate this schema less semantically and more philosophically, we immediately get a semantic contrast-pattern reminiscent of dialectic:
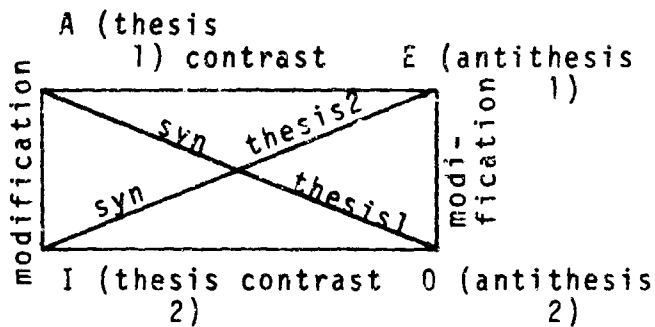
However, if we impose an ordering on this (in order to construct a standard paragraph) we find (as can be seen from the example already given) that we cannot straightforwardly combine $\underline{A}$ and $\underline{O}$ to get synthesis 1 or $\underline{E}$ and $\underline{I}$ to get synthesis 2, for if we could, the paragraph would not progress:

1	Thesis 1	/All+Irish are crooks/	(A)
2	Antithesis 1	/No+Irish are crooks/	(E)
3	Thesis 2	/Some Irish are crooks/	(I)
4	Antithesis 2	/And some Irish+are utterly+non-crooks/	(O)
5	Synthesis	/The Irish go+to extremes:/	
6	I	/they're either+angels or devils/	
7	Synthesis	/Some may+be utter+black+hearted fiends/	
8	II	/but others are absolute+angels of light/	

I should be hard put to it, using the C.L.R.U. model,
to make a machine construct these two syntheses, depending
as they both do on the vital notion of "extreme," which
recapitulates the earlier notion conveyed by "utterly" in
Antithesis 2, i.e., it recapitulates just the part of
Antithesis 2 with is not traditionally part of the propo-
sition O.

I therefore headed this section "The Philosophical Notion
of the Semantic Square"; thereby indicating that the Square of
Opposition, thus linguistically reinterpreted, can only be
used suggestively as a rough guide to fill in the semantic
pattern of a standard paragraph.

With this suggestion in mind, however, let us go back
to the model and its four basic semantic patterns.

The Semantic Square: drawing the second diagonal

It will be seen from the account of the four primary
semantic patterns as given by the model, that not only
intonation and stress, but also position are taken as
being cardinal information-bearers in semantics (semantics
in this being sharply contrastable with syntax). That is
to say, if a semantic match is obtained between two elements,
each in the first position of a template (and therefore each
standing for the first stressed segment of a phrasing) a
different semantic pattern is obtained from that which would
result from a match between, say, the last two elements in
the two templates. Temporal sequence in the one-dimensional
flow of utterance is here projected onto spatial position in
the two-dimensional model; and it is, more than any one thing,
the semantic significance of stressed-position in speech which
is being studied.

Therefore the linguistic reinterpretation of the Square
of Opposition, as set out in the last section, "plays down"
the logical interrelations indicated by the names of the lines
on the Square; it tunes them down to the very lower edge of the

human being's intuitionally perceptible threshold. But it
"tunes up" to a corresponding extent, the actual geometrical
properties of a square, e.g., the fact that a square has four
corners, four equal sides, two equal diagonals.

This raises the question: how on earth can the Square,
consisting of the semantic deep contrast-pattern of a standard
paragraph, be interpreted as having the geometrical properties
of an actual square? How, in particular, can it have four
equal sides, and two equal diagonals, given that in the model,
as just stated above, one-dimensional speech-flow is mapped
onto a two-dimensional spatial frame?

Part of the answer to this question is easy. The "points"
of the square are the stressed "humps" of speech. (34) Spoken
language, even take, at its very crudest, is a string with nodes
in it. Likewise, the equidistance between the points are temporal
equidistances between these main stresses of speech -- at any rate,
in the stressed as opposed to the syllabic languages.(35)

So far, so good. The crunch comes in the question: What
are these diagonals?

To proceed with this, consider again what I asserted
earlier possibly to be the primary overlap pattern of all
language:

/The girl	lived+in+a	house/	GIRL	HOUSE
/and the house	was+in+a	wood/	HOUSE	WOOD
/and the wood	was+full+of	trees/	WOOD	TREES
/and the trees	were+covered+with	leaves/	TREES	LEAVES
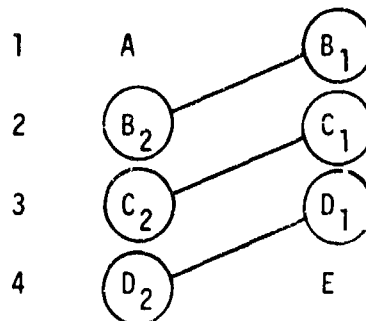
Suppose now that we try to draw in more diagonals.
We find at once that we can draw the diagonals $B_1 -- C_2$
and C_1--D_2: for all we get by doing this is the two pairs
of stressed elements which already occur in the second and
third phrasings, and therefore we know in each case what the
third connecting element is. If we abstract these two
phrasings, moreover, we get quite a sensible pair of actual
phrasings:

1	A	B_1	
2	B_2	C_1	/The <u>house</u> was+in+the <u>wood</u>/
3	C_2	D_1	/The <u>wood</u> was+full+of <u>trees</u>/
4	D_2	E	

The point is that we can't, similarly, draw the other
diagonals, i.e., from $A--C_1$ and from B_2--D_1 because we would
not know how to fill in the phrasings. (Remember, we are not
now doing metamathematically-based referential semantics; we
cannot say that it "follows," by the Transitivity Principle,
that if the girl was in the house and the house was in the
wood, then the girl was in the wood.) For we precisely do not
know whether, in the semantic universe of discourse which the
utterance is creating, it <u>does</u> follow that when the girl was
in the house she was also in the wood. On the contrary, we
don't know yet, but if you ask me for a guess, I should say
it will <u>not</u> follow; if there were bears in the wood, then when
the girl was safe in the house, with the door locked, she
would jolly well <u>not</u> be any longer in the wood; though if
there were also wizards in the wood, as there well might be, who
could come through keyholes and vaporize themselves down
chimneys, then even though she might be in the house and with
the door locked, she would still be (in two more senses of the
phrase) "not out of the wood."

On the other hand, no one is contending that this primary semantic pattern gives us a piece of paragraph; on the contrary, it does not even give us adult discourse.

We get therefore to this thought: <u>perhaps</u> the semantic criterion of the existence of a paragraph -- as opposed to any other indefinitely long sequence of phrasings -- precisely is that in a paragraph we become able to draw the second diagonal. Consider this girl in this wood again. If we compress the sequence not in a syntactic way, by using pronouns, but by using the semantic algorithm which I've just given, which selects the second and third from the sequence of four phrasings:[*] if we do this, we get information about the wood, but we have forgotten the girl. Continue the sequence, however: would it not be very likely to continue (e.g.):

> /The <u>girl</u> was a <u>beauty</u>/
> /Her <u>beauty</u> was <u>dazzling</u>/
> /<u>Dazzling</u> even the <u>very+birds+and+animals</u>/
> /For the <u>very+birds+and+animals</u> knew the <u>girl</u>/
> /That+the <u>girl</u> was a <u>disguised+princess</u>/

If <u>now</u> we try to draw the second diagonal, namely, from the first A to the final element which stands for /<u>disguised+princess</u>/, note that we can; for, applying the algorithm, we shall get out, as a result of this, the final, vital, phrasing (which, note, is also the only phrasing which breaks the monotonous ding-dong pattern of the sequence) which says that the girl was really a disguised princess. And <u>now</u> the sequence of phrasings looks much much more like a paragraph.

So we postulate: finding the paragraph is drawing the second diagonal.

---

[*]Note that to make an intuitively acceptable "abstract" of the sequence, we really want the second and <u>fourth</u> phrasings: to get /the <u>house</u> was in the <u>wood</u>/:/the <u>trees</u> were covered with <u>leaves</u>/, i.e., we have to make use of more <u>intonational</u> features.

The two-phrasing paragraph and the notion of permitted couples

Thoughts of this kind led to the further thought: would it be possible, using the model, to define a minimal paragraph i.e., a paragraph consisting of only two phrasings, within which the machine could discern whether or not there was a semantic square?

Only two types of candidate for such a paragraph intuitively presented themselves:

a) the 2-phrasing double predicate: (Guberina's example)

/<u>Mary</u> milked the <u>cows</u>/
/John did   the <u>goats</u>/

b) the one-phrase question followed by a one-phrase answer (as in an imaginary linguistically condensed Automobile Association phrase-book).

This second form was chosen as the object of study of our first mechanized square-drawing experiment: and the result of it is given in Appendix F.

Using the model to do this experiment, an account of which has been submitted for publication (36), we coded into templates eight short questions and eight short answers. The machine, by doing a semantic match, was required to pair these up so as to produce intelligible discourse, and succeed in doing so, with the exception that the question and answer

/<u>What</u> is the <u>time</u>?/
/<u>Early</u> next <u>week</u>./

could not be eliminated.

In addition to the primary term-anaphora indicated by the match, however, the machine was permitted to discover a secondary semantic connection.

To make this, it first formed permitted couples of all the individual templates; and then looked for other occurrences of these couples as between templates.

For this experiment all permitted couples were taken to be commutative (though the interlingua used for it permitted a term with a slash (a/) to occur in any one position in a template).

Using permitted-coupling on the four primary patterns, it is easy to see that this device greatly increases their semantic interconnectivity, as under:
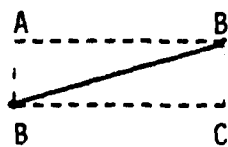
permitted couples

AB
BC
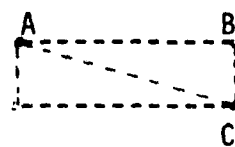
permitted couples

AB
AC
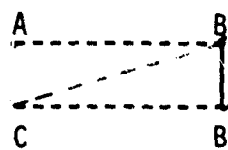
permitted couples

AB
CB cv BC
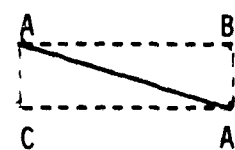
permitted couples

AB cv BA
CA cv AC

If we turn back to our philosophic notion of the Semantic Square for a moment, we see that the notion of permitted couple is standing in both for the notion of minimal semantic contrast

and also for those of <u>reiteration</u> and <u>recapitulation</u>. For
in this program to construct a micro-paragraph, the <u>A</u>, the
<u>I</u>, and the <u>0</u> are to be interpreted only as single terms,
each term standing for one single stressed segment. Synonym,
or anaphora, is indicated by point-name equality: in the
primary semantic pattern E = I, in the second A = I, in the
third E = 0, and in the fourth A = 0. So it is no wonder that
the dialectic pattern vanishes.

In Appendix E the machine output of the experiment is
given. We do not think that it is very good; but it did
teach us to respect the semantic importance of stress-points.

On one linguistic phenomenon it threw considerable light,
namely, on the use of the set of English verbs known as
"anamalous finites". (37) For these are now seen, in at least
one of their properties, as micro-paragraph formers: they
enable the machine to construct the left-hand diagonal.

/<u>Are</u> you <u>coming</u>?/
/<u>Yes</u>, I <u>am</u>./

BE/   MAN: TO:
THIS_R: MAN: BE/

The squaring of an element in a template, as in THIS_R,
indicates a rule, <u>R</u>, of matching-relaxation operating
with regard to it. In this case the rule is: match with
any right hand element of any template (i.e,, draw the right
diagonal) with regard to which a left-diagonal match has
been already achieved.

Schema of the C.L.R.U. Semantic Model

I conclude by giving a schema of the C.L.R.U. semantic model to show that this is a model which, in principle, is mechanizable.

One variant of it is in process of being mechanized by Wilks. (38) But because at this early stage there both are and should be other variants, I give here only an indication of the features which any complete determinate specification of the model would have to cover:

1) Elements

The list of terms, or elements, of the model is as follows:

/INSERT THE ACTUAL FINITE LIST OF N TERMS (N= c. 50) HERE/

2) Connectives

The elements of the model are linked by two connectives, as under:

(i) A colon (:), forming of two isolated terms, a, b, a:b.

(ii) A slash (/), forming of two isolated terms, a, b, a/b.

a:b is commutative; i.e., a:b cv b:a.

a/b is non-commutative.

3) Formulae

A well-formed formula in the model is a pair of formulae linked by a connective and enclosed within a bracket; i.e., (a:b) or (a/b).

Either or both of the formulae so connected can be a single element; i.e., (a/(b:c)).

NOTE: In some variants of the model single-term formulae were also allowed (see Appendix D), these being mentally envisaged as elements connected to null-elements. This was a mistake, as null-elements give rise to far more problems than they are worth. Currently, all the 1-element formulae are being converted to 2-element formulae.

4) <u>Specifiers</u>

In order to give the model more discriminating power, <u>specifiers</u> (i.e., Thesaurus Heads, or Information-Retrieval Descriptors) can be inserted into any formula. e.g., (a MALE:b)

   or (a:b MALE)

   or (a HYDRODYNAMICS/b)

   or (a/b HYDRODYNAMICS)

The set of specifiers used in the model is the following: /INSERT THE ACTUAL LIST OF SPECIFIERS HERE_7

5) <u>Templates</u>

The semantic unit of the system is a template or sequence of three terms of the form(

     a:(b/c)

/LIST HERE ANY OTHER PRIMARY TEMPLATE FORMS WHICH IT IS DESIRED TO PERMIT_7

This is re-expressible as

  αa βb γc,

where the α, β, γ are further specifications, made by using the system, of the stressed words in the original phrasing of which the template in question is to be a coded version. (If more than one template form is permitted, re-express it.)

6) <u>Semantic Match</u>

The unit of operation of the system is a match between two templates. The Rules for semantic matching are as under:

/LIST HERE THE RULES FOR SEMANTIC MATCHING WHICH IT IS DESIRED TO USE_7

7) <u>Semantic Contrast</u>

Secondary semantic connections, or rules of semantic contrast, are also allowed as follows:

/LIST HERE THE RULES OF SECONDARY SEMANTIC CONNECTION WHICH IT IS DESIRED TO USE_7

8)* __Rules of semantic compression for any matched pair of templates__

  /GIVE THESE HERE/

9)* __Recursion-Rules of semantic compression__ (to form the paragraph)

  /GIVE THESE HERE/

10) __Criteria for drawing the left diagonal__ (to test the paragraph)

  /GIVE THESE HERE/

*Sections 8, 9 of this model have not been developed by me but by Yorick Wilks in __Computable Semantic Derivations__.

Notes and References

1. Weinreich, U. "Explorations in Semantic Theory" to appear in Current Trends in Linguistics, Vol. III, Ed. Sebeok.

2. Shillan, D. "A Linguistic Unit Adaptable to Economical Concordance-Making," Cambridge Language Research Unit, mimeo, 1965.
   See also Shillan's earlier book, Spoken English, London, Longmans, 1954, 2nd Ed. 1965.

3. "We neither think nor speak in single words; we express our thoughts in closely-knit groups of words which contribute to the situation in which we are placed at a given moment. Such groups of words are called sense-groups /tone groups7. They are usually separated from each other by pauses, though on occasion these pauses are suppressed...
       "...Their length /i.e., the length of sense-groups, or tone groups7 may vary according to the situation, and the kind of speech being used.... We shall be describing the tunes of English in relation. not to single words or sentences or paragraphs, but to sense-groups."
   O'Connor, J.D. and Arnold, G.F. Intonation of Spoken English, London, Longmans, 1961.

4. These two stress-points are called, respectively, the head and the nucleus; that is to say, the conception of a phrasing which is being used in this paper is that which makes of a tone-group a larger intonational unit, including within itself both a head and a nucleus, and possibly with a lightly-indicated caesura dividing the one from the other (see text 1, Appendix D); not the more restricted conception of a tone-group in which its intonational curve contains only one peak. (In practice a second stress-point is often to be observed as a "silent beat," a phenomenon recognised by phoneticians).
       Both of these notions of tone-group can be defended from the literature, and the apparent discrepancy between the two senses of tone-group turns out to be almost entirely one of terminology, since it is also possible to locate, within the literature, discussion of the differences between major and minor tone-groups.
       "Within units of a certain length /i.e., operation over a major tone-group7 stresses occur at equal intervals of time."

4.   Written notes of two informal talks given by Dr. John
Trim to C.L.R.U. in the Phonetics Laboratory, Cambridge
University, December 19 and 20, 1961. These notes were
checked and corrected by the lecturer, and at the two
talks Prof. Douglas Ellson of the University of Indiana,
and Dr. Bujas of the University of Zagreb were also
present.
       See also Baird, A. "Transformation and Sequence in
Pronunciation Teaching," English Language Teaching,
Vol. 20, 1966, p. 103.
       "In a stress-timed utterance the stressed syllables
tend to occur at equal intervals of time, the intervening
syllables being reduced in prominence."

5.   "...So far /in both the "pictures" of language which
I have given/ I have shown a stretch of utterance with
only one head and one nucleus in it. Moreover, I have
not dealt with the question of how long a stretch of
utterance this schema is meant to show. Is it just a
phrasing, or is it the shortest form of sentence?"
       "Phoneticians of intonational form frequently bemuse
themselves here, because the logical pull of the tradi-
tional sentence is strong upon them. They frequently
talk as though one grammatical sentence could have only
one nucleus, though a moment's reflection would convince
them that this is not the case. I will call such talk
the sentential assumption.
       "Suppose, however, that we do not make the sentential
assumption. Suppose, on the contrary, we assume... [that any
major tone-group] contains two and only two stress points (or
pauses). We will call the first of these the head and the
second the nucleus; longer stretches of utterance will be
considered as being built up of [isochronous] sequences of
these [two]. It then becomes clear that we are taking a much
shorter stretch of utterance than the sentence as our intona-
tional units even though it is normally the sentence, and not
the major tone-group, which is defined by what phoneticians
would call the 'overall intonational tune.'
       "If language is an auditorily-conveyed signal system,
and not just a complex audible outflow from a human being's
lips, the phonetics of intonational form have got to give
the basic auditory mechanism for conveying the signal. And
you have only got to say this for it to become clear that
a sentence is normally far too long a stretch of utterance
to be a single signalling unit.
Margaret Masterman et al. "A Picture of Language," Cambridge
Language Research Unit, mimeo, 1964.

I have quoted myself at length on this subject in order
to try to make clear what abstraction I am making from the
accepted intonational facts in order to postulate the
existence of a semantic unit of speech -- the phrasing --
which is not quite like the ordinary intonational phoneti-
cians' major tone-group, though derived from it. A phrasing
is a major tone group with both a head and a nucleus in it
(see note 3) and with the sequence of heads and nuclei
isochronously spaced throughout the utterance (see note 4)
and which is normally shorter than a sentence (see note 5).

From these three assumptions, taken together, it
follows that there is a two-beat, isochronous ding-dong
rhythm running through all prose, the ding in this being
different from the dong (see also note 10). A further
assumption is then made (see note 5) that this set of
overall intonational facts as here set out constitutes the
primary mechanism for conveying meaning, and is not merely
ancillary to any central analysis of language, i.e.,
"part of stylistics" (see also note 22).

The intonational "picture" given here is, of course,
a simplified first approximation to the facts. In particular,
the interrelationship between major and minor tone-groups is
almost certainly more complex and variable than is here
allowed for; i.e., not all major tone-groups contain one
head and one nucleus; [e.g.] such a group might easily con-
tain two heads, if there were an independent definition of
nucleus and head. On the other hand, the set of facts from
which we have abstracted are acknowledged intonational
facts, which, if we had not first conflated them and then
abstracted from them, might never have been seen in just
this light, i.e., as together constituting possible simple
primary semantic mechanism of language.

6. Gsell, R. et al. "Étude et Réalisation d'un Détecteur de
   Mélodie pour Analyse de la Parole," L'Onde Électrique,
   Vol 43, 1963, p. 556.
   This approach is discussed in: Shillan, D. "A Method and
   a Reason for Tune-Analysis of Language," Cambridge Language
   Research Unit, mimeo, 1965.

7. Dolby, J.L. "On the Classification of Written English
   Phrases," Memo to C.L.R.U., January 1966.
   Dolby, J.L. "On the Complexity of Phrase Translations,"
   Memo to C.L.R.U., January 1966.

8. See the paper by Shillan referred to in Note 2.

9. Literature on Information Retrieval and Machine Translation,
   Bibliography and Index, IBM Research Center, Yorktown Heights,
   New York, 1958.

10. Abercrombie, D. <u>Studies in Phonetics and Linguistics</u>, Oxford, 1965.

11. Regular rhythm as a feature of prose was first noticed by the English poet, Coventry Patmore, in 1856 in his long essay on English metrical law (<u>Amelia</u>, London, 1878). Thereafter it was so completely allowed to lapse from notice that when C.L.R.U. began to utilise it in the summer of 1963 and produced the examples which are given in Appendix D (see C.L.R.U. Lab. Notebook, 1963, Vols. I, III), we thought we had discoved it <u>ab initio</u> -- and indeed by taking it as a two-beat rhythm, perhaps we had.
    Of course, the whole enterprise of reenvisaging tone — groups as breath-groups, and thus of moving over from acoustic to articulatory phonetics (see notes 2 and 35) is bound independently to bring up the question of whether there exists such a rhythm.
    (On the assumptions from which the two-beat rhythm as given here is derived, see also note 5.)

12. <u>The Times</u>, August 13th, 1963.

13. Locke, J. <u>An Essay Concerning Human Understanding</u>, London, 1690.

14. In "A Picture of Language" (see note 5) and at Las Vegas.

15. Margaret Masterman "Semantic Message Detection for Machine Translation, Using an Interlingua," <u>Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis</u> (1961), London, H.M.S.O., 1962, p. 437.

16. Wilks, Y. "Text Searching with Templates" in Masterman, M. et al. "A Picture of Language," Cambridge Language Research Unit, mimeo, 1964.
    Wilks, Y. "Computable Semantic Derivations," Cambridge Language Research Unit, mimeo, 1965.

17. In actual fact the "parts of speech" distinction (for what it is worth) is made more complicatedly, by using sequences of elements in combination.

18. Richens, R.H. "A General Program for Machine Translation between any Two Languages via an Algebraic Interlingua," Cambridge Language Research Unit, mimeo, 1956, abstract in <u>M.T.</u>, Vol. 3, 1956, p. 37.
    Richens, R.H. "Interlingual Machine Translation," The <u>Computer Journal</u>, Vol. 1, 1958, p. 144.
    Sparck Jones, K. "A Note on NUDE," Cambridge Language Research Unit, mimeo, 1963.

19. Allen, W.S., On the Linguistic Study of Languages, Cambridge, 1957.

20. Dixon, R.M.W., What is Language? London, Longmans, 1965.

21. The early history of this suggestion is given in Margaret Masterman, "The Potentialities of a Mechanical Thesaurus," Cambridge Language Research Unit, mimeo, 1956, abstract in M.T. Vol. 3, 1956, p. 36 (read at the International Conference on Machine Translation, M.I.T., 1956).

    This paper itself was never published, owing to mathematical difficulties which arose over the handling of lattices. Since these particular difficulties have now been solved, it is the present intention of its author to prepare it for publication when there shall be time to do so.

    The "thesaurus algorithm" specified in the paper was published as an appendix to a paper entitled "The Analogy between Mechanical Translation and Information Retrieval" (see below). It has actually been used, by applying a theorem in finite lattice-theory which was proved by R.M. Needham, in the C.L.R.U. Information Retrieval System.

Thesaurus Publication of C.L.R.U.

Halliday, M.A.K., "The Linguistic Basis of a Mechanical Thesaurus" Cambridge Language Research Unit, mimeo, 1956, abstract in M.T., Vol. 3, 1956, p. 37.

Margaret Masterman, "The Thesaurus in Syntax and Semantics," M.T., Vol. 4, 1958, p. 35.

Margaret Masterman "What is a Thesaurus?" in Essay on and in Machine Translation, Cambridge Language Research Unit, mimeo, 1959, distributed at the International Conference on Information Processing, Paris, 1959.

Margaret Masterman and Needham, R.M. "Specifications and Sample Operations of a Model Thesaurus," Cambridge Language Research Unit, mimeo, 1960.

Needham, R.M. and Joyce, T., "Thesaurus Approach to Information Retrieval," American Documentation, Vol. 9, 1958, p. 192.

Parker-Rhodes, A.F.. "An Algebraic Thesaurus" Cambridge Language Research Unit, mimeo, 1956, abstract in M.T., Vol. 3, 1956, p. 36.

Parker-Rhodes, A.F. and Wordley, C. "Mechanical Translation by the Thesaurus Method Using Existing Machinery," Journal of the SMPTE, Vol. 68, 1959, p. 236.

Margaret Masterman, Needham, R.M. and Sparck Jones, K. "The Analogy between M_chanical Translation and Library Retrieval," Proceedings of the International Conference on Scientific Information (1958), Washington, D.C. National Academy of Sciences, 1959, p. 917.

Hesse, M.B. "Analogy Structure in a Thesaurus," Cambridge Language Research Unit, mimeo, 1960.

Hesse, M.B. "On Defining Analogy," Proceedings of the Aristotelian Society, 1959-60, p. 79.

Margaret Masterman "Translation," Proceedings of the Aristotelian Society, Supplementary Volume, 1961, p. 169.

22. Guberina, P. Valeur Logique et Valeur Stylistique des Propositions Complexes, Zagreb, Editions Epoha, 1954.

Guberina, P. "La Logique de la Logique et la Logique du Langage," Studia Romanica, 1957.

Guberina, P. "Le Son et le Mouvement dans le Langage," Studia Romanica, 1959.

23. R.M.W. Dixon, op. cit.
See the whole very extensive discussion on the book's central lexical notion, namely, that of an unimaginable vast contextual thesaurus.

24. See also, in the same book, such remarks as:
    "Questions of the form 'is this language?' can only be asked after the description is quite complete (it is potentially so vast that I cannot foresee this ever happening)..." (pp. 105-106)
    See also the cardinal passage (p. 140), "no part of a thesaurus can properly be described until the whole thesaurus is complete"; and, further on down the same page, the passage beginning, "The size of any contextual thesaurus is bound to be enormous..."
    Finally, list the operations which cannot be done, e.g., the determinations of similarities between different sub-classes (p. 147) until the construction of the contextual thesaurus is complete.

25. And Chomsky, up to now, does not utilize the predicate calculus; though possibly he wishes to keep open the possibility of doing so at some later stage.

26. This sentence, as it stands, is agressively worded in a way which I did not intend. It should be re-expressed as:
    "... Fodor and Katz (and even more, Weinreich) start from a philosophic tradition so very different from my own that to me it seems that their dictionaries are always imaginary idealised dictionaries, their examples always

artificially contrived examples, and their problems of
contextual determination unreal problems." The unreality
that I refer to is well illustrated in Weinreich's paper
(see note 1) on p. 137 (note 38), where he says: "We do
not ... propose to hold the semantic theory accountable
for resolving the ambiguity of <u>jack</u> ... in the sentence
<u>I realised we had no jack</u>, by association with, say, <u>car</u>
and <u>break</u> in an adjacent sentence, <u>On a deserted road that
night our car broke down</u>. Such phenomena are in principle
uncoded and are beyond the scope of linguistics, though
they may be both intentional and effective in a "hyper-
semanticised" use of language (Weinreich 1963a:H8).
    Why "hyper-semanticised"? Why not just normal?
Alternatively, are there not considerable grounds for
calling Weinreich's approach to semantics "hyper-syntacticized"
even though he explicitly disclaims this charge on pp. 122
<u>et seq</u>.

27. For the literature on anaphora, see: Olney, J.L. "An
    Investigation of English Discourse Structure with Particular
    Attention to Anaphoric Relationships," Systems Development
    Corporation, Santa Monica, Calif., mimeo, 1964.

28. <u>Freeing the Mind</u>, Articles and Letters from <u>The Times
    Literary Supplement</u> during March-June 1962, <u>The Times</u>
    <u>Publishing Co.</u>, 1962.
    Mary Hesse "Analogy Structure in a Thesaurus" (see Note 21).

29. Shillan was for more than twelve years Director of a School
    of Languages. My judgment underlying the whole "intonational-
    semantic" approach of this paper is that the practical
    teachers of spoken language, spurred on both by the urgent
    pressures of their professions and by their continual,
    intimate contact with the experimental material, have thought
    about language more simply, more unselfconsciously, more
    fundamentally, and so, in the end, more scientifically than
    so-called "paralinguistic scientists."
        I would myself always go to ɩ practising spoken-language
    teacher to learn basic facts about intonation and to a
    practising simultaneous translator to learn basic facts about
    translation: but to say this, of course, is to declare a
    personal bias.

30. There is some very beautiful and pure scholarship on what
    Aristotelian logic really was to Aristotle to be found in
    the first two chapters of: Lukasiewicz, J. <u>Aristotle's</u>
    <u>Syllogistic</u>, Oxford, 1951.

31. Margaret Masterman et al. "Semantic Basis of Communication,"
    Cambridge Language Research Unit, mimeo, 1964.

32. Margaret Masterman, "The Semantic Basis of Human Communi-
    cation," Arena, No. 19, April 1964, p. 18.

    Margaret Masterman, "Commentary on the Guberina Hypothesis,"
    Methodos, Vol. 15, 1963, p. 139.

33. Margaret Masterman, "A Picture of Language" (see note 5).

34. We start with a representation of a continuous piece of
    spoken language.

```
    A_____B_____C_____D
    .           .          .           .
```

The lines AB, CD are phrasings and the points A,B,C,D
stress-points (isochronous beats). The broken line BC
is a pause for breath or break. We think we have empirical
evidence that phrasings occur in pairs, so the above is the
lowest level or simplest unit of language. We can measure
AB,CD (i.e., the time interval between stress-points) on a
Gsell machine. Although there will be slight variations
according to the speaker, we can regard these values as
constant. BC (i.e., how long the break is) can clearly vary
much more, and if BC is variable, so is AD. So we say that
there are two kinds of information that we can gather from
spoken language -- constant and variable. We now re-repre-
sent the flow of speech as a square in order to illustrate
what we can learn form these different kinds of information.
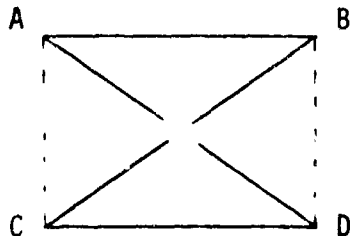
AB,CD are fixed and from the basis of the square BC,AD
are variable; thus the break can occur anywhere within
certain limits; but these lengths are determined by AB,
CD and are thus represented as diagonals.
AC,BD are inferred connections.
This is in a sense "new" information which we immediately
see when we represent language in this way.

35. Stetson, R. Motor Phonetics, Amsterdam, 1951.

36. Dobson, J. "Report of an Experiment to find Semantic Squares"
    in an Interlingually-Coded Text taken from a Travellers'
    Handbook, Cambridge Language Research Unit, mimeo, 1965.

37. Shillan, D. "Anamalous Finites," Cambridge Language Research Unit, mimeo, 1965.

38. Wilks, Y. "Computable Semantic Derivations" (see Note 16).

APPENDIX A

Comparative Quatrain Analyses of English and French

(See p. 4 of text)

(For the notion of a Quatrain, see the text, pp. 7 et seq)

Below is given an analysis of a text in Canadian English
and its translation into Canadian French. The phrasings were
marked by hand. The main stresses are here marked with double
underlining, otherwise the notation.

1 In a review ( )————————→1 en passant en revue

2 of existing homemaker+programs—→2 les programs+actuels des
soins+a domicile

3 and other community services——→3 et les+autres services+
communautaires

4 for+the elderly, ( )————— →4 [destinés] aux vieillards

- -

5 it+is noted ( )———————→5 on remarque ( )

6 that+there+are encouraging——→6 un progres encourageant
developments+in+Canada

7 although services are not+so+   7 [a cet+effet ( )]
extensive

8 [as they are] in some+other+   8 au Canada. ( )
countries

9 bien+que les services
[disponibles]

10 ne soient pas+encore

11 aussi+vastes ( )

12 que dans certain autres+pays

- -

9 where+the needs of+the   13 ou pour certaines raisons
elderly

10 for services at community+   14 on a accorde plus+d'attention
level

11 have, for various reasons,   15 aux besoins des personnes+
agees

12 received greater emphasis   16 a l'échelon local.

- -

phrasings or sub-phrasings inserted by the tr iator, in
whole or in part, to restore the balance of the prose.

mapping of the translation-correspondence between phrasings,
from the English to the French.

1 As+a <u>matter</u> of <u>fact</u>————————>En <u>fait</u> ( )

2 I+was <u>struck</u> by+the <u>fact</u>————→j'ai été <u>etonne</u> de <u>constater</u>

3 that+there+are <u>so</u> <u>few</u>————————→qu'il <u>existe</u> si <u>peu</u>

4 <u>visiting homemaker+services</u>————→de <u>services+de+soins+à+domicile</u>
⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯[au Canada]

- -

5 In <u>this</u> report————————————— >Dans <u>notre</u> <u>mémoire</u>

6 our <u>estimate</u> <u>is</u> ——————————— >nous avons <u>mentionné</u> <u>l'existence</u>

7 that+there+are <u>fifty</u> <u>five</u> ——— >d'environs <u>cinquante-cinq</u>

8 <u>visiting+homemaker+services</u>——— >services de <u>soins+à+domicile</u>
[in Canada]

- -

9 The <u>Red Cross</u>————————————→9 La <u>Croix+Rouge</u> [<u>Canadienne</u>]

10 operates <u>thirty</u> of <u>these</u>——→10 en <u>gère</u> <u>trente</u>.

11 ( ) ( )————————→11 ( ) ( )

12 ( ) ( )————————→12 ( ) ( )

- -

13 There+are other organisations→Il existe <u>d'autres</u> <u>organismes</u>

14 Providing these services————————→qui <u>fournissent</u> ces <u>services</u>,

15 such+as <u>visiting+homemaker+</u>————→tels les <u>Associations+d'Aides+</u>
<u>associations</u> ( ) <u>familialies+visiteuses</u>

16 <u>family+service+agencies</u> ( )——→les <u>bureaux</u> <u>d'aide+aux+familles</u>

- -

17 <u>children's+aid+societies</u> ( );→les <u>sociétes</u> <u>d'aide+aux+enfants</u>

18 the <u>V.+O.+N.</u>————————————→les <u>infirmières+visiteuses</u> de
<u>l'ordre+de+Victoria</u>,

19 and <u>some+others</u>.——————————— >et <u>quelques</u> <u>autres</u>.

20 ( ) ( )——————————> ( ) ( )

- -

21 These <u>services</u> ( )————————— > <u>Ces</u> services

22 are <u>extremely</u> <u>important</u>————→sont <u>vitaux</u> ( )

23 for <u>elderly</u> <u>people</u>.————————→ pour les <u>personnes</u> <u>agées</u>

24 ( ) ( )———————— · ( ) ( )

- -

25 By <u>assisting</u> in <u>household</u>+   25 En+les <u>aidant</u> à <u>accomplir</u>
   <u>tasks</u>

26 and <u>personal</u>+<u>services</u> (   )   26 leurs <u>tâches</u> <u>menagères</u>

27 they <u>help</u> [<u>many</u>]+of+<u>them</u>   27 et en+leur <u>vendant</u> (   )

28 [<u>to live independently</u>],   28 des <u>services</u> <u>personnelles</u>,

- -

29 and they <u>postpone</u> (   ),   29 ces <u>organismes</u> <u>retardent</u> souvent

30 and in <u>some</u>+<u>cases</u> make   30 la <u>necessité</u> des soins dans
   <u>unnecessary</u>   <u>les hôpitaux</u>

31 the <u>need</u>   (   )   31 ou même les rendent

32 for <u>institutional</u> <u>care</u>.   32 <u>quelquefois</u> <u>inutiles</u>

- -

```
1 It is recognised (  ) ──────────→1 Il est reconnu (  )
2 that+the building of schools ──→2 que la construction d'écoles
3 and+the expansion of programs ──→3 et l'expansion des programmes
4 alone ───────────────────→ 4    (  )      (  )
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5 are not the complete+answer ──→5 ne sauraient resoudre
6 to+the training problem ──→ 6 a+elles seules
7  (  )       (  )      ──→7 tout le probleme
8  (  )       (  )      ──→ 8 de la formation. (  )
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

APPENDIX B

Printout of phrasings in the pilot scheme of the C.L.R.U.
Semantic Concordance SEMCO [SEMANTIC CONCORDANCE]

(cf. p. 4 of text)

Key

1) number immediately above each phrasing is its text-position
   no.
2) 8 marks first stress-point in a phrasing (which is written
   on the left): e.g.   8   THINK
3) 6 marks second stress-point in a phrasing (which is written
   on the right): e.g.   6   QUESTIONS
4) the pendant (consisting of the unstressed words) is printed
   under the first stressed word and within brackets; 6 stands
   for an opening bracket, 9 for a closing bracket. This, in
   the first phrasing (10101), the pendant-components are:

          6 1 9 6 THAT 9 6 9

   which may be re-expressed

           (I)   (THAT)   ( )
5) Silent beats are also indicated by the figure-sequence
   6   9, i.e., by brackets.
   e.g. (phrasing 10201), /in+a review   (7/

SEMCO | 8 REVIEW   ( )
ENTRY | (IN A) ( ) ( )

```
      10101
o  T...                        o  ...

o  I  96  THAT  ...  9

      10102
o  MORE IMPORTANT             o  QUESTIONS

o  OF  THE  96  96  9

      10103
o  COMMUNITYDEVELOPMENT       o  CANADA

o  ARISING FROM 96  IN  96  ..
      10104
o  HO.                        o  MARSHAL

o  IS  96  WE  CAN  96  9

      10105
o  HEALTH                     o  WELFARE

o  OUR  96  AND  96  SERVICES9

      10106
o  LOCAL                      o  LEVEL

o  AT  THE  96  96  9

      10107
o  PROVIDE                    6  SERVICE

o  TO  96  THE  KIND  OF  96  9

      10108
o  PREVENTIVE                 6  POINTOF VIEW

o  THAT  FROM  A  96  96  TELL  9
      10109
o  HELP                       6  PEOPLE

o  96  96  9

      10110
o  REMAIN                     o  LONGASPOSSIBLE

o  TO  96  AS  96  9
      10111
o  NORMAL                     6  ACCOMODATION

o  IN  THEIRSO  96  96
      10112
o  NORMAL                     6  SETTING

o  AND  THEIR  96  96  ..
      10201
o  REVIEW                     6  9

o  IN  A  96  96  9
      10202
o  EXISTING                   o  HOMEMAKER PROG
RAMS
o  OF  96  96  9
      10203
o  COMMUNITY SERVICES         6  ELDERLY

o  AND  OTHER  96  FOR  THE  96  9
      10204
o  NOTED                      6  9

o  IT  IS  96  96  9
      10205
o  ENCOURAGING                o  DEVELOPMENTS

o  THAT  THERE  ARE  96  96  IN  CANADA  9
      10206
o  SERVICES                   6  EXXTENSIVE

o  ALTHOUGH96  ARE  NOT  AS  96  9
      10207
o  ARE                        6  OTHERCOUNTRIES

o  AS  THEY  96  IN  SO  96  9
      10208
o  NEEDS                      6  AGED

o  WHERE  THE  96  OF  THE  96  9
      10209
o  SERVICE                    6  COMMUNITY LEVE
L
o  FOR  96  AT  96  9
      10210
o  VARIOUS                    6  REASONS

o  HAVE  FOR96  96  9
      10211
o  GREATER                    o  EMPHASIS

o  RECIEVE96  96  9
      10212
o  9                          6  9
```

```
o  yo  yo  o
   2010?
o  MATTER              6 FACT

o  AS A 9b OF 9o 9
   20102
o  STRUCK              6 FACT

o  I WAS 9b BY THE 9b 9
   20103
o  SO                  6 FEW

o  THAT THERE ARE 9b 9b 9
   20104
o  VISITING            6 HOMEMAKER SERV
ICES
o  9b 9b 9

   20201
o  THIS RESPECT        6 ESTIMATE IS

o  IN 9b OUR 9b 9
   20202
o  FIFTY               6 FIVE

o  THAT THERE ARE 9b 9b 9
   20203
o  VISITING HOMESERVICES o CANADA

o  9b IN 9b 9
   20301
o  RED CROSS           6 THIRTYOF THESE

o  THE 9b OPERATESSO S
   20401
o  OTHER               6 ORGANISATIONS

o  THERE ARE 9o 9b 9
   20402
o  PROVIDING           6 SERVICES

o  9b THESE 9o 9
   20403
o  VISITINGHO EASSNS    6 9

o  SUCH AS 9o 9o 9
   20404
o  FAMILYSERVICESAGENCIESS o

o  9b So 9
   20405
o  CHILDRENSAIDASS S    o 9

o  9b 9b 9
   20406
o  THE VON              o 9

o  9o 9b 9
   20407
o  OTHERS               o 9

o  AND SO ESO 9o 9
   20408
o  9                    6 9

o  9o 9b 9
   20501
o  SERVICES             o IMPORTANT

o  THESE 9b ARE EXTRE ELY 9o
   20502
o  ELDERLY              6 PEOPLE

o  FOR 9b 9b 9
   20601
o  ASSISTING            6 HOUSEHOLD TASK
S
o  BY 9b IN 9b 9
   20602
o  PERSONAL SERVICES    o 9

o  AND 9o 9b 9
   20603
o  HELP                 6 MANY OF THEM

o  THEY 9b 9b 9
   20604
o  LIVE                 6 INDEPENDENTLY

o  TO 9b 9b 9
   20605
o  POSTPONE             6 9

o  AND THEY9o 9o 9
   20606
o  SUCH CASES           6 UNNECESSARY

o  AND IN 9b MAKE 9b 9
   20607
o  NEED FOR             6 INSTITUTIONALC
ARE
o  THE 9o 9o 9
```

20102
o STRUCK   6 FACT

o I WAS 96 BY THE 96 9

    10209
o SERVICE   6 COMMUNITY LEVE
L
o FOR 96 AT 96 9
    10206
o SERVICES   6 EXXTENSIVE

o ALTHOUGH96 ARE NOT AS 96 9
    20301
o SERVICES   6 IMPORTANT

o THESE 96 ARE EXTREMELY 96 9
    20103
o SO   6 FEW

o THAT THERE ARE 96 96 9
    20606
o SOME CASES   6 UNNECESSARY

o AND IN 96 MAKE SO 9
    20406
o IN YOU   6 9

o 96 96 9
    20201
o THIS RESPECT   6 ESTIMATE IS

o IN 96 OUR 96 9
    10101
o THINK   6 ONE

o I 96 THAT 96 9
    20601
o ASSISTING   6 HOUSEHOLD TASK
S
o BY 96 IN 96 9
    10207
o ARE   6 OTHERCOUNTRIES

o AS THEY DO IN SOME96 9
    20405
o CHILDRE SAID96 S   6 9

o 96 96 9
    10209
o COMMUNITY SERVICES   6 ELDERLY

o AND OTHER 96 FOR THE 96 9
    10103
o COMMUNITYDEVELOP.E T 6 CANADA

o ARISING FROM 96 IN 96 9
    20502
o ELDERLY   6 PEOPLE

o FOR 96 96 9
    10202
o EXISTING   6 HOMEMAKER PROG
RAMS
o OF 96 96 9
    10205
o ENCOURAGING   6 DEVELOPMENTS

o THAT THERE ARE 96 96 IN CANADA 9
    20404
o FAMILYSERVICESAGENCIES96 9

o 96 96 9
    20202
o FIFTY   6 FIVE

o THAT THERE ARE 96 96 9
    10211
o GREATER   6 EMPHASIS

o RECIEVED96 96 9
    10105
o HEALTH   6 WELFARE

o OUR 96 AND 96 SERVICES9
    10109
o HELP   6 PEOPLE

o 96 96 9
    20603
o HELP   6 MANY OF THEM

o THEY 96 96 9
    10104
o HOW   6 MARSHAL

```
• IS 96 BE CAN 96 9
   20604
• LIVE                    6 INDEPENDENTLY

• TO 96 96 9
   10106
• LOCAL                   6 LEVEL

6 AT THE 96 96 9
   20401
• OTHER                   6 ORGANISATIONS

• THERE ARE 96 96 9
   20407
• OTHERS                  6 9

• AND SOME96 96 9
   20602
• PERSONAL SERVICES       6 9

• AND 96 96 9
   20605
• POSTPONE                6 9

• AND THEY96 96 9
   10108
• PREVENTIVE              • POINTOF VIEW

• THAT FROM A 96 96 WILL 9
   20402
• PROVIDING               6 SERVICES

• 96 THESE 96 9
   10107
• PROVIDE                 6 SERVICE

• TO 96 THE KIND OF 96 9
   20301
• RED CROSS               6 THIRTYOF THESE

• THE 96 OPERATES96 9
   10201
• REVIEW                  6 9

• IN A 96 96 9
   10110
• REMAIN                  6 LONGASPOSSIBLE

• TO 96 AS 96 9
   10210
• VARIOUS                 6 REASONS

• HAVE FOR96 96 9
   20403
• VISITINGHOMEASSNS       6 9

• SUCH AS 96 96 9
   20104
• VISITING                6 HOMEMAKER SERV
ICES
• 96 96 9
   20203
• VISITING HOMESERVICES 6 CANADA

• 96 IN 96 9
   20101
• MATTER                  6 FACT

• AS A 96 OF 96 9
   10102
• MOREIMPORTANT           6 QUESTIONS

• OF THE 96 96 9
   20607
• NEED FOR                6 INSTITUTIONALC
ARE
• THE 96 96 9
   10208
• NEEDS                   6 AGED

• WHERE THE 96 OF THE 96 9
   10204
• NOTED                   6 9

• IT IS 96 96 9
   10111
• NORMAL                  6 ACCOMODATION

• IN THEIR96 96 9
   10109
• NORMAL                  6 SETTING

• AND THEIR 96 96 9
   20408
• 9                       6 9

• 96 96 9
   10212
• 9                       6 9

• 96 96 9
```

```
6  SERVICE
6  TU BE THE . N  JF 46 .
   20402
6  PROVIDING              6  SERVICES

6  96 THESE 96 9
   20501
6  RED CROSS             6  THIRTYOF THESE

6  THE 96 OPERATES 96 9
   10111
6  NORMAL                6  ACCOMMODATION

6  IN THEIR 96 96 9
   10208
6  NEEDS                 6  AGED

6  WHERE THE 56 OF THE 96 9
   10103
6  COMMUNITY DEVELOPMENT  6  CANADA

6  ARISING FROM 96 IN 96 9
   20203
6  VISITING HOMESERVICES 6 CANADA

6  96 IN 96 9
   10209
6  SERVICE               6  COMMUNITY LEVEL

6  FOR 96 AT 96 9
   10205
6  ENCOURAGING           6  DEVELOPMENTS

6  THAT THERE ARE 96 WA IN CANADA 9
   20201
6  THIS RESPECT          6  ESTIMATE IS

6  IN 96 OUR 96 9
   10205
6  COMMUNITY SERVICES    6  ELDERLY

6  AND OTHER 96 FOR THE 96 9
   10206
6  SERVICES              6  EXXTENSIVE

6  ALTHOUGH 96 ARE  UT AS 96 9
   10111
6  GREATER               6  EMPHASIS

6  RECIEVE 96 96 .
   20102
6  STRUCK                6  FACT

6  I WAS 56 BY THE 96 9
   20101
6  MATTER                6  FACT

6  AS A 96 OF 96 9
   20103
6  50                    6  FEA

6  THAT THERE ARE 96 96 9
   20202
6  FIFTY                 6  FIVE

6  THAT THERE ARE 96 96 9
   20601
6  ASSISTING             6  HOUSEHOLD TASKS

6  BY 96 IN 96 9
   20104
6  VISITING              6  HOMEMAKER SERVICES

6  96 96 9
   10202
6  EXISTING              6  HOMEMAKER PROGRAMS

6  OF 96 96 9
   20501
6  SERVICES              6  IMPORTANT

6  THESE 96 ARE EXTREMELY 96 9
   20607
6  NEED FOR              6  INSTITUTIONALC ARE

6  THE 96 96 9
   20604
6  LIVE                  6  INDEPENDENTLY

6  TU 96 96 9
   10106
6  LOCAL                 6  LEVEL
```

```
● A   THE  YE  SE Y
    10203
● RELAY                        ● LONGESPOSSIBLE

● TO  YE  AS  YEY
    10207
● ARE                          6 OTHERCOUNTRIES

● AS  THEY  YE  IN  SOMEYE  Y
    20401
● OTHER                        6 ORGANISATIONS

● THERE  ARE  YE  YE  9
    10101
● THINK                        6 ONE

● I  YE  THAT  YE  9
    20502
● ELDERLY                      6 PEOPLE

● FOR  YE  YE  Y
    10109
● HELP                         6 PEOPLE

● YE  YE  9
    10108
● PREVENTIVE                   ● POINTOF VIEW

● THAT  FROM  A  YE  YE  WILL  9
    10102
● MOREIMPORTANT                6 QUESTIONS

● OF  THE  YE  YE  9
    10210
● VARIOUS                      ● REASONS

● HAVE  FORYE  YE  Y
    20600
● SOME  CASES                  ● UNNECESSARY

● AND  IN  YE  MAKE  YO  Y
    10105
● HEALTH                       ● WELFARE

● OUR  YE  AND  YE  SERVICESY
    10104
● HOW                          ● MARSHAL

● IS  YE  WE  CAN  YE  Y
    20603
● HELP                         ● MANY OF THEM

-  THEY  YE  YE  Y
    10212
● 9                            ● 9

● YE  YE  9
    20406
● THE  VON                     ● 9

● YE  YE  Y
    20403
● VISITINGORGANIZATIONS        ● 9

● SUCH  AS  YE  YE  Y
    10201
● REVIEW                       ● 9

● IN  A  YE  YE  Y
    20605
● POSTPONE                     ● 9

● AND  THEYYE  YE  Y
    20602
● PERSONAL  SERVICES           ● 9

● AND  YE  YE  Y
    20407
● OTHERS                       6 9

● AND  SOMEYE  YE  Y
    10204
● NOTED                        ● 9

● IT  IS  YE  YE
    20404
● FACILITYSERVICESAGENCIESYE 9

● 3E  YE  Y
    20408
● 9                            ● 9

● YE  YE  9
    20405
● CHILDRENSAIDASSNS            ● 9

● YE  YE  9
```

APPENDIX

From "A Note on Finding Phrasings in Raw Natural Language Text by Algorithm" by John Dotson.

... The basic information on which the algorithm depends is of two types:

a) **Syntactic.** We do not need a complete parsing program, nor yet a complete syntactic theory. What we do need to be able to spot is entities that most syntax programs can spot, e.g., conjunctions (but not the limits of conjunct groups), prepositional phrases (but not their qualificands), etc. More detail of syntax requirements will be found in the statement of the algorithm.

b) <u>Temporal or quai-syllabic.</u> To each word we attach a number which represents, broadly speaking, the amount of time it takes to say it. This number may correspond to the number of syllables in the word, but does not necessarily do so. For example, to the word "characteristics" with 5 (char-ac-ter-ist-ics) syllables we attach the number 4, as the first three syllables come out faster than the last two.

The output of the algorithm is, as already stated, the boundaries of the phrasings, which we call bar-lines or bars; but

one refinement to this simple scheme needs to be noted before
the algorithm can be given: that of the splittable phrasing
or semi-bar.

Frequently we find that a phrasing is too long for it to
be coded up by 3 elements in the interlingua NUDE or NUB
without serious loss of message, and yet that the phrasing certainly
corresponds to one breath group. We also find that 2 consecutive
phrasings (breath groups) are both very short and that a triple
more properly corresponds to the concatenation of the two
phrasings. To deal with the first such case, we split up the
phrasing (the place of the partition being determined algorithmic-
ally) by a semi-bar, and represent each of the halves by 3 elements
(some of which may be null), but yet treat the whole phrasing as
the unit for squaring purposes. We may term this a "divorce."
Correspondingly, a "marriage" occurs when two successive phras-
ings are short and the bar that divides them is attenuated to a
semi-bar.

The rules for dividing up the phrasings can now be given.

Section I   To find the bar lines

    1) Put bar lines after any punctuation and after
        the closing bracket of a noun group or preposi-
        tional phrase or adverbial subjunct.
        EXCEPTION: If two commas separate precisely
                one word, delete the first comma.

    2) Put a bar before the last monosyllabic word of
        a group (of 1 or more) monosyllabic words. NB.
        This rule is not recursive.
        EXCEPTION: Do not split a noun group except
                before a conjunction.

    3) Any bars occurring after

>   1) a conjunction
>   ii) to, used infinitely   } call these
>   iii) a nominal subjunction   } special words
>   iv) a preposition   }

are moved one word to the left.

4) If a bar consisting of a single note has been
created under 2) (as modified by 3)), delete it.

Section II   To change full bar lines to semi-bar lines

5) A punctuation bar occurring within a noun group
is attenuated to a semi-bar

6) A bar whose temporal value is < 5 has its closing
bar relegated _if_ the following word is special;
otherwise, its opening bar is relegated _unless_ the opening
bar is a punctuation, in which case nothing
happens.

The results of this simple algorithm are quite good (see
appendix, which contains a text phrased by hand following the
algorithm), but there is every likelihood of their being improved
by further considerations based on the number of non-special
words in any bar, for we find that most of the clear errors that
remain are of some bars being of excessive length. Further,
consideration has been given to the possibility of detecting
the main and subsidiary stressed words in each bar, based on
the following observations.

i) Special words are never stressed.

ii) The first word in each bar is not stressed unless
the bar is very short.

Reference

D. Sheillan. "A Method and a Reason for Tune Analysis." C.L.R.U.,
M.L. 179.

Although there has been,/in recent years,/some retardation/
in the pace'of Canadian aircraft development,/the national statis-
tics/show a continuing increase/in the airborne activity'of the
country/both in terms'of public air transportation/and in business
and private aviation./ Whereas in earlier years/the frontiers'of
aircraft development'in Canada/tended'to reflect/the military
need'for high speed flight/ the facts'of defence policy/and of
the aircraft market'generally/have deflected'the Canadian
aircraft industry/towards sophisticated,'relatively low-speed
aircraft/having unique performance characteristics,/which will
compete favourably/with foreign designs./ The effect'of these
industrial trends/on the program'of the division/has been'to
emphasise the work/on design and development problems/of vertical'
and short take-off aircraft,/and on various aspects/of flight
safety/and utilisation./

In support'of industrial design'and development,/the work/
in the division's wind tunnels/has been primarily devoted/to
aerodynamic investigations/of new designs'of aircraft and rockets,/
and to certain non-aeronautical problems/of ships superstructures/
and structural space-frame members./ At the same time,/the
division's structures laboratory/completed/the structural
development'and proving/of a new light aircraft,/carried out/the
ground vibration analysis'for it,/and cleared the aircraft'for
flutter/in a comprehensive series/of flight investigations./ Of
more basic interest,/an inflight evaluation/was made'of the control-
ability requirements/of vertical take-off aircraft/by means'of a
variable-stability helicopter/developed'in the division's flight
research laboratory./ It is gratifying to note/that these
programs/were all undertaken/at the request of industry,/ and with
the continuing cooperation/of industrial representatives./

Although/in principle all/of the aeronautical work/of the
flight research laboratory/is in some way/related to flight and
safety,/certain of its projects/are more directly concerned/with
accident avoidance or mitigation./ A crash position indicator,/
developed in recent years,/is now in commercial production./
It was, however,/originally intended for use/in subsonic aircraft,/
and arising from a desire/to exploit the device/on supersonic
military aircraft,/it has been necessary/to do a great deal of
research/on its supersonic deployment characteristics./ These have
now been shown/to be admissible/and full-scale trials are pending./
Other contributions/to the flight safety area of the work/have
involved a study of aircraft crash dynamics,/and continuing support
and evaluation/of quality control procedures,/and scientific
support/of aircraft crash investigations/including/a very heavy
involvement/in the study/ of the Montreal disaster/of November
1963./

Concerning aircraft utilization,/ the division's efforts/have
been directed/towards those areas of national activity/where aerial
methods/might offer economies in cost,/or improvements in effect-
iveness./ These include agricultural applications,/forest fire
fighting,/aerial logging,/high sensitivity magnetic surveys,/
precipitation physics,/and studies of atmospheric turbulence./

During the year, also,/the basic research/of the Division/
gave rise/to a number of papers/on swirling flow,/hypersonic aerody-
namics,/flow separation,/the aerodynamics of bluff bodies,/and
fatigue of materials./

APPENDIX D

<u>Examples of the Algorithm of removing the primary
and secondary stresses from texts and then trying to guess
at the message from these alone</u>. (See, in text p. 11)

The sequence of sets of stresses is given first, and
then the sequence of texts, correspondingly numbered.

1.

		2.	
PUT	HANDS	WHEN	BOY
BEEN	DIVERSION	HAD	CLOCK
SOME	IDLE	PENDULUM	( )
HEAVY	HOURS	LIFTED	OFF.
IF	BEEN	FOUND	CLOCK
GOOD+LUCK	PROVE+SO	VERY+MUCH	FASTER
ANY	THINE,	WITHOUT	PENDULUM.
( )	( )	( )	( )
THOU	HALF	IF	PURPOSE
PLEASURE	READING	CLOCK	GO,
I+HAD	WRITING+IT,	CLOCK	BETTER
( )	( )	LOSING	PENDULUM.
THOU	LITTLE	TRUE,	NO+LONGER
THINK	MONEY	TELL	TIME,
I+DO	PAINS	THAT	NOT+MATTER
ILL	BESTOWED.	TEACH+ONESELF	INDIFFERENT
		PASSAGE	TIME.
		LINGUISTIC	PHILOSOPHY,
		ONLY	LANGUAGE,
		NOT	WORLD ,
		BOY	PREFERRED
		CLOCK	WITHOUT+THE+PENDULUM.
		ALTHOUGH	NO+LONGER
		TOLD	TIME,
		MORE+EASILY	BEFORE,
		MORE+EXHILARATING	PACE.

3.

SOON	SLUMP		
FOLLOWED	BOOM		
UNIONS	AGAIN+FIGHTING		
PRESERVE	STANDARD+OF+LIVING.		

TRIPLE+ALLIANCE COLLAPSED
BLACK+FRIDAY 1921,
RAILWAYMEN TRANSPORT+WORKERS
WITHDREW DECISION
SUPPORT MINERS'+DISPUTE,
( ) ( )

BECAUSE ALLEGED
MINERS DID+NOT+CONSULT
( ) NEGOTIATIONS
( ) ( ).

MINERS BEATEN
AFTER LONG+STRUGGLE,
ENGINEERS DOWN
FOLLOWING YEAR.

INDUSTRY INDUSTRY
WAGES REDUCED.
( ) ( )
( ) ( )

4.

BLACK+SILK NECK-CLOTHS
ALWAYS AVERSION.

IT SIGNAL
( ) DESPAIR,
SIGN END
SIGHT. ( )
AFTER+THAT EVERYTHING
( ) SUPPORTED+HIM
KEPT+HIM IN+BEING,
( )
DISSOLVED VANISHED.
SELF-RESPECT ( )
( ) ANYONE
DINE BILL.
PAY

5.

NOW	APPARENT
PARTICULAR	NAME
NOT	SAME+MEANING
THROUGHOUT	PROGRAM.
THIS	PARTICULARLY+USEFUL
CASE	LABELS.
USE	NESTING+BLOCKS
ECONOMISES	STORAGE+SPACE,
SPACE	OBTAINED
LOCAL	VARIABLES
BLOCK	BLOCK+IS+ENTERED
RELINQUISHED	BLOCK+IS+LEFT.
ALTHOUGH	BEGINNER
TEMPTED	DEFINE
HEAD	OUTER→BLOCK
ALL	VARIABLES
USED	PROGRAM,
BETTER	
AS	DEFINE+VARIABLES
	REQUIRED.

6.

AS+WHEN	TRANCÉD
SUMMER	NIGHT,
GREEN+ROBED	SENATORS
MIGHTY	WOODS,
TALL+OAKS	BRANCH+CHARMED
EARNEST	STARS,
DREAM,	SO+DREAM
ALL+NIGHT	WITHOUT+A+STIR,
SAVE	ONE+GRADUAL
SOLITARY	GUST,
COMES	SILENCE
DIES	OFF,
AS+IF	EBBING+AIR
JUST	ONE+WAVE,
SO+CAME	THESE+WORDS,
( )	AND+WENT...

7.

THIS	JANET,	HERE	ISAAC+NEWTON,
THIS	JOHN,	GREAT+MAN	SCIENCE.
THIS	MOTHER,	NEWTON	HAD
THIS	FATHER.	GREAT	MIND.
		UNDER	APPLE+TREE.
SEE	JANET,	( )	( )
MOTHER,	( )	THOSE	APPLES
SEE	JANET	OVER	HEAD
PLAY	( ).	APPLE	ON
		BRANCH	TREE.
THIS	FATHER.	APPLE	OFF
SEE,	JOHN+AND+FATHER.	BRANCH.	( )
SEE	DOG,	CAME	DOWN.
JANET,	( )	( )	( )
SEE	LITTLE	CAME	DOWN
DOG	( ).	NEWTON'S	HEAD.
COME,	LITTLE+DOG,	BLOW	APPLE
COME,	JANET.	GAVE	NEWTON'S+HEAD
SEE	LITTLE+DOG	GAVE	IDEA
PLAY	( ).	NEWTON.	( )
		MADE	QUESTION
		COME+INTO	NEWTON'S+MIND.

TEXT 7

(N.B. Since a special study was made of this passage, it
is given as phonetically annotated by Shillan. Note also
the single <u>caesura</u>, or cut, within each phrasing.)

I
1   I+have'put/in thy`hands
2   What has'been/the diversion
3   of'some/of+my`idle
4   and'heavy/ hours.

II
1   'If/it has`been
2   the'good+luck/to`prove+so
3   of'any/of thine,
4   ()   ()

III
1   and thou/hast but`half
2   so much'pleasure/in`reading
3   as'I+had/in writing+it,
4   ()   ()

IV
1   'thou wilt/as`little
2   'think/thy money
3   as'I+do/my`pains
4   'ill/be`stowed.

John Locke, <u>Essay Concerning Human Understanding</u>. Preface,
First Edition, 1690. Edition used: Oxford University Press, 1894.

2)
I
1   /When I+was a <u>boy</u>/
2   /I+had+a clock/
3   /with+a pendulum()/
4   /which+could+be <u>lifted off</u>./

II
1 /I <u>found</u> that+the clock/
2 /went <u>very+much</u> <u>faster</u>/
3 /<u>without</u> the <u>pendulum</u>/
4 /() ()/

III
1 /<u>If</u> the main <u>purpose</u>/
2 /of+a <u>clock</u> is+to <u>go</u>,/
3 /the <u>clock</u> was the <u>better</u>/
4 /for <u>losing</u> its <u>pendulum</u>./

IV
1 /<u>True</u>, it+could <u>no+longer</u>/
2 /<u>tell</u> the <u>time</u>,/
3 /but <u>that</u> didn't <u>matter</u>/
4 /if one+could <u>teach+oneself</u> to+be <u>indifferent</u>/
5 /to+the <u>passage</u> of <u>time</u>./

V
1 /The linguistic <u>philosophy</u>,/
2 /which+cares <u>only</u> about <u>language</u>,/
3 /and <u>not</u> about+the <u>world</u>,/
4 /is+like+the <u>boy</u> who <u>preferred</u>/
5 the <u>clock</u> <u>without+the+pendulum</u>./

/because, <u>although</u> it <u>no+longer</u>/
/ <u>told</u> the <u>time</u>,/
/it+went <u>more+easily</u> than <u>before</u>/
/and+at+a <u>more+exhilarating</u> <u>pace</u>./

Ernest Geffner, <u>Words</u> <u>and</u> <u>Things</u>. (Introduction by Bertrand Russell, p. 15)

3) 1 /Soon slump/
I   2 /followed the boom/
    3 /and+the unions were again+fighting/
    4 to preserve their standard+of+living./

    1 /The Triple+Alliance collapsed/
    2 /on "Black+Friday" in 1921,/
II 3 /when+the railwayman and transport+workers/
    4 /   withdrew their decision/
    5 /to support a+miners' dispute,/
    6   ()   ()

    1 /because they alleged/
III 2 /the miners did+not+consult them/
    3 / () in+the negotiations./
    4 / ()   ()   /

    1 /The miners were beaten/
IV 2 /after a long+struggle,/
    3 /and+the engineers went down/
    4 /the following year./

    1 /In industry after industry/
V   2 /wages were reduced./
    3 / ()   ()   /
    4 / ()   ()   /

Eric L. Wigham, Trade Unions. (Home University Library, p. 36)

4) 1 /Black+silk neck-cloths/
    2 /had always been his aversion./
I   3 /() It+was+a signal/
    4 /() of   despair,/
    5 /a sign that+the end/
    6 /was+in sight.   () /

II
1 /After+that, everything/
2 /()that+had supported+him/
3 /and kept+him in+being./
4 dissolved.   ()

III
1 /His self+respect vanished./
2 /   ()   ()   /
3 /He+would dine with anyone/
4 Who+would pay the bill./

Virginia Woolfe, "On Beau Brummel." (The Second Common Reader)

5)

I
1 /It+is now apparent/
2 /that+a particular name/
3 /may not have+the same+meaning/
4 /throughout a program./

II
1 /This is particularly+useful/
2 /in+the case of labels./
3 /The use of nesting+blocks/
4 /also economises storage+space,/
5 /since space is obtained/
6 /for+the local variables/
7 /of+a block when+the block+is+entered/
8 /and+is relinquished when+the block+is+left./

III
1 /Although the beginner/
2 /may+be tempted to define/
3 /at+the head of+the outer+block/
4 /all the variables/
5 /used in+the program,/
6 /it+is better to define+variables/
7 /as they+are required./

6)

I   1 {   /As+when, upon a tranced/
    2   /Summer night,/

II   1 {   /These green+robed senators/
     2   /of mighty woods/

III   1 {   /Tall+oaks, branch+charmed/
     2   by+the earnest stars,/

IV   1 {   /Dream, and so+dream/
    2   /all+night without+a+stir,/

V   1 {   /Save from one+gradual/
   2   /solitary gust,/

VI   1 {   /Which comes upon+the silence/
    2   /and dies off,/

VII   1 {   /As+if the ebbing+air/
     2   /had just one+wave,/

VIII   1 {   /So+came these+words,/
      2   / ()   and+went../

John Keats, Hyperian.

7)   1 {   /This is Janet/
I   2   /This is John/
   3   /This is Mother/
   4   /This is Father/

   1   See Janet,
   2   Mother (),
II   3   See Janet
   4   Play. ()

III 1 { <u>This</u> is <u>Father</u>.
2 { <u>See</u> John+and+<u>Father</u>.

IV 1 { <u>See</u> the <u>dog</u>,
2 { Janet, ( )
3 { <u>See</u> the <u>little</u>
4 { <u>dog</u>. ( )

V 1 { <u>Come</u>, <u>little+dog</u>,
2 { <u>Come</u> to <u>Janet</u>,
3 { <u>See</u> the <u>little+dog</u>
4 { <u>play</u>. ( ).

Mabel O'Donnell and Rona Munro. (illustrated by Florence
and Margaret Hoppes), "<u>Off to Play</u>," <u>The Janet and John
Book</u>. (Nesbit and Co.)

8) I 1 { /<u>Here</u> is Sir <u>Isaac+Newton</u>,/
2 { /the <u>great+man</u> of <u>science</u>./
3 { /<u>Newton</u> had/
4 { /a <u>great</u> <u>mind</u>./

II 1 { /<u>He+is</u> <u>under</u> an <u>apple+tree</u>/
2 { / ( ) ( )/
3 { /<u>Those</u> are <u>apples</u>/
4 { /<u>Which+are</u> <u>over</u> his <u>head</u>.

III 1 { /The <u>apple</u> was <u>on</u>/
2 { /a <u>branch</u> of+the <u>tree</u>./
3 { /The <u>apple</u> <u>came+off</u>/
4 { /the <u>branch</u>. ( )/

IV 1 { /It <u>came</u> <u>down</u>/
2 { / ( ) ( )/
3 { /It <u>came</u> <u>down</u>/
4 { /on <u>Newton's</u> <u>head</u>./

```
        1 ⎧ /The blow which+the apple/
   V    2 ⎪ /gave to Newton's+head/
        3 ⎨ /gave an idea/
        4 ⎩ /to Newton. ()/


  VI    1 ⎧ It made a question/
        2 ⎨ /come+into Newton's+mind./
```

I. A. Richards and Molly Gibson. A "Basic English" Text
from *English Through Pictures*..

APPENDIX E

Extract from Computer Printout of an Interlingua Sample
(See text, p. 20)

This sample has been randomized for the purpose of mechanically analysing it by transforming it from Italian to English word-order. It should be interpreted from the <u>right</u>, therefore, not the <u>left</u>. Thus, the interpretation of the first entry should be, "That use of the Italian root <u>fond</u> (if any) which is entry number 1991 in this sample and which means the same as some use of the English word <u>shareholding</u>."

Key

The figure 8 stands for the interlingual connection, <u>:</u> (colon).

The figure 7 stands for the interlingual connective, <u>/</u> (slash).

The letter N stands for NOT.

The number-sequence 6-9 stands for opening and closing brackets, <u>( )</u>.

Thus the interlingual formula for the first entry can be re-expressed:

$$(CAUSE/((SELF:PAIR)/HAVE)):SIGN$$

```
SHAKEHOLDING   6CAUSE766SELFHPAIN97HAVE9O8SIGN
SHAKP          6MUCHBCANY866PARTHTHING9B6RAIN9
SHARPHITTED    BANGBTHINK
SHATTER        BANGB6CAUSE76BE7NWHOLE99
SHAVE          CAUSE76NHAVE76PARTH9AN BEARD99
SHEETOFPAPER   PARTHSTUFF PAPER
SHEETOFPAPER   ONEHSTUFF PAPER
SHINE          CAUSE76MUCHBSENSE SEE9
SHINING        6CAUSEBSENSE SEE9BHOW
SHIRT          666MANBTHING97IN9USTUFF98PART
SHFFT          CAUSE76BEBLOHEY
SHOP           6FOLK76CAUSE766SELFUPAIN97HAVE9996PARTBHFNO
SHORT          6HAVE7NMUCHBLINE99BHOR
SHORT          NMUCHB6PUTBHEN9
SHORT          NMUCHBLINE
SNOW           6CAUSE76HVE7SIGN99BPOINT
SNOW           PUINT76FOR76CAUSE7SENSE999
SHOWN          DONEB6CAUSE7SENSE SEE9
SNUT           CAUSE7CHOYOHE76INBHAVE99
SNUT           CAUSE7DHOIOHE76INDHAVE99
SNUT           CAUSE7DHOIOHE76INBHAVE99
SNUT           CAUSE7DHOIOHE76INDHAVE99
SIDE           PARTBHERE
SIDE           HAVEB6PARTBHHERE9
SIDEPLACK      PART
SIGN           SIGN
SIGN           CAUSE7BHAVE76SELFBSIGN99
SIGNATURE      SELFBSIGN
SILVER         STUFFB
SIMILAR        SAMEBBE
SING           CAUSE7UCAUSE76BE76RAIN MUSIC99
SINGER         6CAUSE76UCAUSE76BE76RAIN MUSIC999BUAN
SINGING        6CAUSE76UCAUSE76BE76RAIN MUSIC999BHOY
SISTERINLAW     6CAUSE76BE7YAN9BLINE9BUAN
SITTING        BHEBMAN9BHOW
SITTING        MAN7DO
SITUATED       PUINTBHERE
SITUATED       HAVEBHERE
SKILFULNESS    6MUCHBCANY9BHAVE
SKIN           6B6MANBTHINGS7IN9USTUFF9HPART
SKIRT          6CHANGEBSHEREBULINE
SKY            UPBHORLD
SLEEP          6CHANGE7HEY96661A57HE9BHOB9
SLEEP          6BEBMAI9BHOB
SLEEP          MAN7DO
SLEEP          MAN7BE
PUTTOSLEEP     CAUSE76HAN7HEY9BHOB
SLENDER        6HAINBHHERE
SLICE          PART
SLOWLY         HAVE76HUCD9BPARTHHMEN9
SMASH          BANGB6CAUSE7NBFY
SMASH          BANGB6CAUSE7NWHOLEY
SMITH          6BFOR76UCAUSE76BE7THING9U99HAUSE7STUFF7ROBY9HAY
SMOKE          CAUSE76HEATBSTUFF9
SMOKE          HEATBSTUFF
SMOKE          CAUSE76HEATBSTUFF9 SMOKE
SMOKING        6CAUSE76HEATBSTUFF9HHOY
SO             FOR
SOCIETY        FOLKB6RAIN
SOCIETY        6PARTBFOLK9
SOFT           6SENSE FEELS9HHOY
SOME           6NUCH9BCOUNT
SON            6B6CAUSE76BE76ANYBLINE9BUAN
SON            6B6CAUSE76BE77YANYBLINE9BUAN
SONG           6CAUSE76BE76RAIN MUSIC99BKIND
ASSOONAS       PUINTBHEN
SORT           KIND
SOTHAT         CAUSEB6ANTBLINEB9IGN9
SOTHAT         FOR76LINEBUSIHNY
SOTHAT         6CAUSEB6ANT97BLIYEHSIGN9
SOTHAT         CAUSE7BANTBLINEB9IGN9
SOTHAT         FOR
SORROWFUL      FEELBHPLEASE
SOUL           6INBMANY9B6HAVEBTHING9
SOUL           TRUEBSELF
SOURCE         6CAUSE76BEYOTHING
SPARKLING      CAUSE76UUCHBSENSE SEEY
SPIN           6CHANGE7IHFHE9B66BHEBPARTBHHERE99
SPIN           6CHANGEB/SHEREY9BHAIN
SPINDLE        PARTB66CANT7IE9BTHING9
SPLIT          6CAUSE7NYHOLEUBLINE
SPOIL          CAUSE76BEBHPLEASE9
SPOIL          CAUSE76CHANGEBHPLEASE9
SPOIL          CAUSE76HAVE7BNHPLEASEB6RAIN99
SPOILT         DONEB6HAVE76NHPLEASEB6RAIN99
SPOILT         DONEB6CAUSE76HEBHPLEASE99
SPOON          6HANTUSE9BTHING
SPRING         6CAUSE76BE7STUFF LIQUID9H6BPARTBHHERF9
SPEAKER        6CANT9HUCHB6CAUSE76HAVE73IGN99
TALK           CAUSE76HAVE76THINGBBSIGN99
TALK           THINGBSIGN
TALK           THINGBSIGN
STANDITARE     6CHANGE7THIN9
STANDING       HAVEB6HAIN
STANDING       DOBHOB
STANDING       HAVEB76PARTBHHEREY
STAR           6UPBHORLD9BTHINGH
STEP           UPBHOB
STEEPPATH      6PUNT76CHANGE7HREH9B6UPBLINE9
STICK          6HANTUSE9BTHING
```

1991 FONG
1992 AGU77
1993 AGU22
1994 FRANTUM
1995 BARR
1996 FOGU
1997 FOGLI
1998 BRILL
1999 FOLGAR
2000 CAVIC
2001 GERVOGLI
2002 BOTTFR
2003 BREV
2004 BREV
2005 BREV
2006 DIMOSTP
2007 DIMOSTH
2008 ESPOST
2009 CHIND
2010 CHIUS
2011 CHIUS
2012 CHIUS
2013 CANT
2014 FIANC
2015 FIDUC
2016 CENN
2017 FIRM
2018 FIRM
2019 ARGENT
2020 CONFOR
2021 CANT
2022 CANTATOO
2023 CANT
2024 COGNAT
2025 ASSIS
2026 ASSIS
2027 GIAC
2028 GIAC
2029 BRAVUR
2030 GONNELL
2031 ELUO
2032 CIEL
2033 ADDORMENT
2034 ADDORMENT
2035 DORM
2036 DORM
2037 ADDORMENT
2038 GRACIL
2039 FETT
2040 ADAGIO
2041 FRACASS
2042 FRACASS
2043 FANNP
2044 FUM
2045 FUM
2046 FUM
2047 FUMANT
2048 COSI
2049 EHT
2050 EXT
2051 BLAND
2052 ALCUN
2053 FISLUOL
2054 FIRLI
2055 CANN
2056 APPENA
2057 EMEH
2058 AFFICCHE
2059 AFFINCHE
2060 AFFINCHE
2061 ACCIO
2062 ACCIO
2063 DOLOR
2064 ANIM
2065 ANIM
2066 FONT
2067 FRIZZANT
2068 FIL
2069 FIL
2070 FUS
2071 FENDITUR
2072 GUAST
2073 GUAST
2074 GUAST
2075 GUAST
2076 GUAST
2077 CUCCHIA
2078 FONT
2079 SALOUT
2080 TRANS
2081 TRANS
2082 FRAGLING
2083 FONT
2084 ESEN
2085 SESETT
2086 SESETT
2087 ASTO
2088 ERT
2089 ERT
2090 BAST

STILL	NDO	2091FERM
STILL	HAVE76MUCHb6PARTbWHEN99	2092ANCORA
STILL	DOWG5AH6bLINE9	2093ANCORA
STINKING	CAUSE7b6SENSE SMELLHHPLEASE9	2094FERTIL
STUCREACHAH6FULK7oCAUSE76b6EIFHPAIR97HAVE99HRPARTbWHERE9		2095b09b
STUCKING	b6bMANoTHING97IN9b3TUFF9bPART	2096CAL7
STOP	NMOREbHOW	2097FERM
STOP	NMORE	2098FERM
STOP	6CHANGE7HHERE9bNbOHE	2099FERM
STURH	6bANb6OUPb6PARTBDOHLU999ADO	2100BUIDASC
STURY	6bDUHEECHANGE9bCDUNT9bSIGN	2101CONT
STUVE	6CAUSE7HEAT9b6bMAN7USE9bTHING9	2102FORM
STRAIGHT	THUEbLINE	2103DIRITT
STRAIGHT	TRUEbLINE	2104DRITT
STRANGER	6HSELFoFOLK9bHAN	2105FORESTIFF
STRAWBERRY	PARTbPLANT FRUIT	2106FRAGOL
STREAM	LINEbSTUFF LIQUID	2107CORRENT
STRENGTH	b6UCHbCAN7DC99bHOH	2108FORZ
STRENGTH	b6UCHbHE9bHOH	2109FORZ
STRIKE	MUCHbDU	2110COLP
STRIKE	HANbbDU	2111COLP
STRING	LINEbSTUFF	2112FILZ
STRING	b6IN7USE9bTHING97FOR7CAUSE76REbPART99	2113FILZ
STRING	6bAN7USE9b6THINGHLINE9	2114FILZ
STRINGY	6HUCHbLINE9bHAVE	2115FITOS
STRIVE	WILL7DU	2116CERC
STRIVE	WILL7DU	2117CERC
STRIVE	WILL7DU	2118CERC
STROKE	6HHUCHb6bHANGbDO99bHOH	2119COLP
STROKE	6bANb6DU9bHOH	2120COLP
STRONG	MUCHbBE	2121FORT
STRONG	b6UCHb6CAN7DU9	2122FORT
STRONG	MUCHb6CAN7DU5	2123FFRR
STRONG	MUCHb6CAN7DO9	2124GAGLIARD
STRONGEST	MUCHb6UCHSTRONGEST	2125FORT
STRONGEST	MUCHb6D UCHb6CAN7DO99	2126FORT
STRONGHOLD	6HAVE7b6UCHb3TUFF99766HAN7IN9R6THINGbWHERE99	2127CASTELL
STRONGHOLD	6b6UCHbHAVE97STUFF9b6PARTHHHERE"	2128CASTELL
STRONGLY	b6UCHbHOH	2129FORTE
STRONGLY	MUCHb6bCAN7DU9	2130FORTE
STRONGLY	MUCHb6bCAN7DU5	2131FORTEMENTE
STRONGLY	MUCHbHU9	2132FORTEMENTE
STUDY	6bHAN7IH9b6THINGb7HERE99HPART	2133GARIN
STUFFED	HAVE7ob6UCHbSTUFFb57IN9	2134GREH
SUBMIT	BILL7bHAVE76CHINGHEbNPLEASE99	2135CHIH
SUBMIT	6FOR7bPLEASE7HbSEIF99b6CAUSE76SELF7DO99	2136CHIH
SUBSTANCE	STUFF	2137CORH
SUBTLE	b6UCHb6COUNTHHPART99bHAVE	2138FIN
SUBTLE	MUCHb6HAIH	2139FIN
SUBURB	b6PARTbHHERE9b6b6IUCHbFOLKHRE99HPART	2140DORG
SUBWAY	6FOR7oCHANGE7HHEHE9bHb6b1AN7IN9HRTHINGbHHEREbHbbPART	2141SAUFZ
SUCCESS	DUNEbPLEASE	2142FORTUN
SUFFICIENCY	b6UUCHIFLEASFbbHAVE9	2143ABBASTH ZA
SUFFICIENCY	6FOR7Ub6HHAVE	2144ABBASTA ZA
SUFFICIENCY	HbANT	2145ABBASTI ZA
SUFFICIENT	b6UCHIFLEHbSE9bHAVE	2146BASTHNT
SUFFICIENT	HbANT	2147BASTANT
SUFFICIENT	6FOR7Ub6E9bHAVE	2148BASTAHT
SUDDEN	COUNTb6PARTbHMENS	2149ESTAT
SUMMIT	UPbHU2HRE	2150CIM
SUMMIT	6UPbHU DHE9bHHERE	2151CIH
SUMMIT	UPb6PARTHHHEHE9	2152CIH
SUPPER	b6UHEbLMA CHbFULK	2153CEN
SUPPER	b6CAUSE7b6HHAVF7LIFEHHISTUFFbHRbPARTbHHENb	2154CEN
SUPPLY	CAUSE7HAVE	2155FORN
SUPPORT	b6CAUSE7UbHTbCHANGE7HHERE9bHTHING6	2156APPOGR
SUPPORT	6CAUSE7bTHING7UP9bHTHINGH	2157APPOGG
SUPREME	CAUSE7bSELF7HHAVF7bTHUFFbISIGN999	2158ASSICUH
SURGEON	6CAUSE7b6HULFHLIFEY99bHAN	2159CHIRURG
SURE	6b6UCMbCOUNTbHHASTY:7HIb7SSA EbbPARTH HEHF999	2160FOK ICH
SURE	b6UCHb6COUNTHHEASTY9bHHHbIHSAHF9	2161FONHICH
SURE	CAUSE7b6HAVE7b6HUIbITHTIHEbBSIHH9	2162SIUR
SUREL	SENSEbPLEASE	2163DOLC
SURELINESS	6SENSEbUPLEASFb9HHUR	2164DOLC
SURELINESS	6CAUSE7bFEELbHPLEASE9bHHIH	2165DOLC
SURELY	b6CAUSE76bSEIFb6PbI797HbVENbHHHTHHEbHHO6	2166FROH
SURELY	6FOR7bCAUSE7HSEIF7HAVEbHHHFbDO3HTRUEH	2167DOH
SURLIER	b6UCHb6UHEYHHE	2168ANHEI
TACITURN	HbUCHHCHbUSFbHAVE7S1GH bH	2169CHIH
TAIL	PARTbbEAST	2170COD
TAIL	COUNTbPART	2171COD
TAIL	PARTbbLINE	2172COD
TABLED	6bHANGE7UPb97HbEEIbHPLFAHE9	2173HAFFEZIO
TASK	6HCANbHbAHTYbDU	2174CARIC
TEA	COUNT	2175DIECI
THAT	PUINT	2176CIO
THATIS	6PUIHTbbAbFb9IF	2177CIOF
TILE	PUIHT	2178IL
THEFT	CAUSE7bSEIF7bHAVF7bHbSEIFbHTHINGbH99	2179FUHT
THEN	PUIHTb6HH9	2180ALLORH
THERE	PAHTH bFbH	2181CI
THESE	PUIHTbHHbHF	2182C
THEYARE	PUIHTbbCUUHTbHbANH	2183COLHI
THICK	b6UCHbSTUFFbHHCH	2184FITTY

THICK	6=UCH6=COUNT=THING9976IH76SAME=7PART=WHERF9=9	2|R5FITY
THICKLY	6=UCH6=COUNT=THIN=997IN76SAVE6=PART=WHERE9=9	2|R6FITYAM= TP
IHING	THING	2|R7ARTICOL
IHING	THING	2|8RCO9
IHING	THING6	2|R9ARTICOL
IHIS	POINT	2|90CIO
THIS	POINT=nSELF=WHERE9	2|91CIO
THREAD	LINE=STUFF	2|92FIL
THREAD	6=AN7USE996THING6LINE9	2|99FIL
THRIVE	6=UCH=LIFE96=HAVE	2|94FI9=
THRIVING	6=UCH=LIFE96=HAVE	2|95FI0P
THROAT	PART=BEAST	2|9=30L
THROUGH	IN=LINE6	2|97ATT=AV= 30
THRU=	CAUSE7uTHING76CHANGE7=ERE99	2|9R8FTT
THRUST	8ANG6uCAUSE7IN9	1|99FICC
TICKET	THING=SIGN	2200R16LIFT
TIRED	CRANGE/=HE66FEELL=IPLEASE=9	2201AMNOI
=E TIRED	N=UCH=nCAN7DU9	2702ANNOI
TO	=HERE6	220=A
TU	IN6=HERE	2204A
TOAST	FOLK=DU	22=59RI=DI=
TOE	PART6=AN	2206DIT
TOIL	6=UCH=DU96=nPLEASE	2207FATIC
TOIL	8UCH=DUU=n=PLEASE=	2208FATIC
TU=URROU=	PART6=HER	22n9DnHAH|
TUNURHU=u	PART6=HER	2210DnMATTI =
TOOLS	COUNT6uu=AN7USE96THING9	2211FFH=
TOOTH	PART6=BEAST	2212DENT
TO=H	6=PART6=HEHE 96=UCH=66=AN7IN9=6THING6=WHFHE9=9	2913CITTA
TO=N	6=PART6=HEHF9=66=UUCH=FOLK9=RE9	2914CITTA
TOY	6=CAUSE7=FEEL=nPLEASE99=THING	2915=I0=ATT =
TREASURY	6=HAVE=cSIGN=STUFF9=9=FOLK9=9=PART6=HERF9	291=FI=F
TREE	THEE =PLANT	2917ALR=P
TRIAL	=FUR/=PUI=T7=6FULK=n=AIN9=9=IRH9966DDNFOLK9	291=6|UN|=
TRICKING	6=PLEASE=6=nAIN9=THING	291=9VA=H
TRIP	=FUR7=uE7=nLEASE=nPART=nHENF9=99=6CHAN=E7=HFRF9	22=0RIT
TROUBLE	=UONE6CHANRE9=nPLEASE	22=1Y|
TROUBLE	DU=E9=nPLEASE	2222AUAI
TROUBLE	DU=n=PLEASE	229=FATIC
TROUBLE	6=UONE6CHAN=nE96DuPLEASE	2924FATIC
TAKE=TROUBLE	6=UCH=D=c=F=6=7=CAUSE7=DE7THIN=9=9=	2925DISTU6=
TROUBLESUM=	6=CAUSE7=FEEL=nPLEASE99=MD=	2926FA=TI=I =
TROUSERS	66=MAn=THIn=97IN9=STU=F9=	2927CAL?0=
TRUNK	6=AN7USE7=THING.	22=8=UL
TRUST	=THIn=/THUE9=HD=	8229FIDA=Z
TRUST	CAUSE7=uPU|=THUA=97=HAVE7SELF99	2=30CO=FI=
TRUSTED	=PU=T=UA=976HAVE77THIN9=	2231FIDAT
TRUSTY	=PU=T= A=97CA=76HAVE7THIN=9=	2=32FIDAT
TURN	6CHAN=E7=HER=6nRSAMEHLINE9	2=330DD
TURN	CHANGE7=66CHAN=E7 HERE9=nLINE=	22=4=IR
TURN	6CHAN=6ELINE96HU=	2235=IR
TURN	66CHA=LE7=E=E9=8=n=AMERLINE9=HU=	2=36=IR
TURN=LINE	6CHAN=E7=HER=9=n=SAME=LINE9	2=70IVFRT
TURN=LINE	CHAN=E7=6CHA=6E7=HEn9	2R=9DIVFHT
TO=N	PART66 ALL=PAIN9	22=98EHFLL
TO=N	PART=PAIN	2240GFnELL
TO=	COUNT	22410U=
TOLY	CAUSE7=nPLEA=E=SE=SE=	2242PHUTT
=G=F=L	CAUSE/=A=7=HAVE/=nPLEASF=nIRN=9=	2=44DIS=0=T =
=nG=F=L	CAUSE/==A=7=7=FEEL=nPLEASE99	2=44DIS=UN= =
=nVERN=TAn=	HAVE7=nTn=n=nU29=8IGN9	2=44C=P
=NUER=T=nG	THU=nTH=K	2=44C=P
=nUER=TA=6	CAUSE/ n=VF7=n=I6=66=ILL7RD=9=	2=47A=9=N=
=nUER=TA=6	CAUSE/=n=VF7=n=I6=66=ILL7RD=9=	2=4A=9=N=
=nUER=TA=6	CAUSE/=n=A=F7=n=I6=66=ILL7RD=9=	2=4=9=N=
=nDULATE	CHAn=E=UP	2=5=FLUTT=
=REUUAL	N=A=E=n=	2=S|DI=n=U=
=REUUAL	N=A=E=C=UN=	2=53DI=Cn=
=n=A=UUM=AGLE	n=A=T/=IMIN=THF9	2=5=CONTHA=
=n=IT	=n=6THI=n	2=5=ELEM=NT
=n=IT	=H=	2=59RELE=n=
=nLI=L	N=A=E	2=5=DI9=|=|
=nUCCUPIED	N=U	2=57DI=O=C= A=
=nPLEA=ANT	nPLLA=E=H=9	2=5=DI=n|A= =C
=n=RI=HT	UP=RU= UP=I=HT	0=R9ERE=T
=n=RI=HT	UP=HU=	2=4RENE=T
=A=E	=U=FUL=	2=4|C09=UH
=5E	U=E=	2=4=AnO=E=
=E=9=U=U	=U76=A=L=6=UU=FOLF=9	2=4=A=C=9=
=TU=E=	66CHAn=L7H=Y7HAVF976n09=FOL==	2=4=A=IT=
=SEFUL	CAN7=nC7UDF9	2=5=A=OV
=SUAL	6A=E=B=	2=4=A=IT=A=
=UAL	=UU=B=c=	2=7A=IT=A=

APPENDIX F

Output of the "A.A. Phrasebook" experiment
(see text, p. 32)

Key

A unit of computer output consists of
   i) a question
  ii) an answer with which the question matched
 iii) information about the pattern of match
         The first line of information represents the
      template corresponding to the question, and the
      second line represents the template corresponding
      to the answer. The letters O represent the first
      and third elements of the templates, and have been
      separated by a vertical manuscript bar. A direct
      match is indicated by the occurrence of the letters
      A or C following the matching element; a permitted
      couple by the letters X and/or Y following the
      matching elements.

   In order to clarify the output, manuscript diagrams have
been added. In these, the horizontal lines indicate the
assumed semantic connection between the first and third elements
of the template, a double vertical or diagonal line indicates
a direct match, and a dashed horixontal or vertical line indi-
cites a permitted couple.

The phrases and their codings

Q1 WHERE   DOES   HE   LIVE
                WHERE8   MAN8   IN7

Q2 ARE   YOU   ILL
                UNPLEASE6N SELF8   FEEL7

Q3 WHEN   DO   YOU   GO
                WHEN6N SELF8 CHANGE7

Q4 WHERE   IS   THE   TOILET
                WHERE8 THING8   IN7

Q5 WHERE   ARE   YOU   GOING
                WHERE8N SELF8   TO7

Q6 WHAT   IS   THE   TIME
                WHEN8   BE7   WHEN8

Q7 WHY   ARENT   YOU   DRESSED
                WHY8N SELF8   WRAP7

Q8 WHICH   IS   THE   WAY
                WHERE8   BE 7   LINE8

Q9 WHERE   DO   I   EAT
                WHERE8   SELF 8   IN7

Q10 WHERE   CAN   I   GO
                WHERE8   SELF 8   TO7

Q11 WHEN   DOES   IT   LEAVE
                WHEN8 THING8   FROM7

Q12 WHEN   ARE   YOU   COMING
                WHEN8 WHERE8 CHANGE7

A1 QUARTER PAST   FOUR
                WHEN8   BE7   POINT8

A2 AT   THE   TRANSPORCAFE
                IN7   KIND8   POINT8

A3 TO   THE   NEXT   VILLAGE
                TO7N SAME8 POINT8

A4 I'M   VERY   WELL   THANKYOU
                PLEASE8   SELF8   FEEL7

A5 AT   THE   TERMINUS
                IN7   TO8   POINT8

A6 DOWN   THAT   STREET
                IN7   THIS8   LINE8
                TO7   THIS8   LINE8

A7 I   THOUGHT I   WAS
                THINK7   SELF8 HIMIS8N

A8 EARLY   NEXT   WEEK
                WE7N   SAME8   WHEN8

A9 I   DONT   KNOW
                h TRUE7   SELF8 NTHIS8N

A10 JUST   HERE
                THIS8   BE7   POINT8

Q1	WHERE	...ES	DO/GE	LIVE	...	IN7
Q2	ARE	YOU	ILL	APLEASE8N SELF8		FEEL7
Q3	WILL	DO	YOU	GO	WHERE8N SELF8	CHANGE7
Q4	WHERE	IS	THE	TOILET	WHERE8 THING8	IN7
Q5	WHERE	ARE	YOU	GOING	WHEN8ON SELF8	TO7
Q6	WHAT	IS	THE	TIME	WHEN8	BE7	WHEN8
Q7	WHY	ARENT	YOU	DRESSED	WHY8N SELF8	WRAP7
Q8	WHICH	IS	THE	WAY	WHERE8 BE 7	LINE8
Q9	WHERE	DO	I	EAT	WHERE8 SELF 8	IN7
Q10	WHERE	CAN	I	GO	WHERE8 SELF 8	TO7
Q11	WHEN	DOES	IT	LEAVE	WHEN8 THING8	FROM7
Q12	WHEN	ARE	YOU	COMING	WHEN8 WHERE8 CHANGE7	

A1	QUARTER PAST FOUR	WHEN8	8E7	POINT8
A2	AT THE TRANSPORCAFE	IN7	KIND8	POINT8
A3	TO THE NEXT VILLAGE	TO7N	SAME8	POINT8
A4	IM VERY WELL THANKYOU	PLEASE8	SELF8	FEEL7
A5	AT THE TERMINUS	IN7	TO8	POINT8
A6	DOWN THAT STREET	IN7	THIS8	LINE8
		TO7	THIS8	LINE8
A7	I THOUGHT I WAS	THINK7	SELF8	6THIS98
A8	EARLY NEXT WEEK	BE7N	SAME8	WHEN8
A9	I DONT KNOW N TRUE7	SELF8	6THIS98	
A10	JUST HERE	6THIS98	8E7	POINT8

LIST OF PERMITTED COUPLES, FOLLOWED BY THE NUMBER OF OCCURRENCES OF EACH

```
BE7       WHEN8      04
IN        LINE8      04
IN7       POINT8     06
PLEASE8   FEEL7      01
THINK7    6THIS98    02
WHEN8     CHANGE7    06
WHEN8     POINT8     06
WHEN8     WHEN8      04
WHEN8     FROM7      04
WHERE8    IN7        15
WHERE8    LINE8      04
WHERE8    TO7        08
6THIS98   POINT8     07
TO7       LINE8      03
TU7       POINT8     02
WHY8      WHAP7      02
NPLEASE8  FEEL7      02
N  TRUE7  6THIS98    03
                     12
```

Q1 WHERE DOES HE   LIVE TRANSPORCAFE
A2   AT   THE

O X |OA Y
OAX |O

Q1 WHERE DOES HE   LIVE TERMINUS
A3   AT   THE

O X |OA Y
OAX |O

Q1 WHERE DOES HE   LIVE STREET
A4   DOWN THAT

O X |OA Y
OAX |O

Q1 WHERE DOES HE   LIVE KNOW
A5   I   DONT

O |OC
O |OC

WHERE DOES HE LIVE
JUST HERE

OA O Y

ARE YOU ILL
I'M VERY WELL

OA O Y

THANKYOU
O YOC

ARE YOU ILL
I THOUGHT I

O OC Y

WAS
OC O

ARE YOU ILL
I DON'T KNOW

O OC

O OC

Q1 ^10
Q2 ^4
Q2 ^7
Q2 ^9
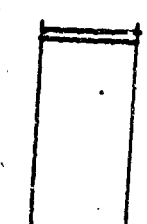
Q3 A1 — WHEN DO YOU GO / QUARTER PAST FOUR
OAXY|O
OAX|O Y

Q3 A4 — WHEN DO YOU GO / EARLY NEXT WEEK
OCXY|O XY
O X|OC Y

Q3 A3 — WHEN DO YOU GO / I DONT KNOW
O |OC
O |OC

Q4 A2 — WHERE IS THE THE TOILET TRANSPORÇAFE
O X|OA Y
OAX|O Y

Q4 A4 — WHERE IS THE THE TOILET TERMINUS
O X|OA Y
OAX|O Y

WHERE	IS	THE	TOILET	
	DOWN	THAT	STREET	O X Y OA Y
				OAX O
WHERE	IS	THE	TOILET	
		DONT	KNOW	O OC
				O OC
WHERE	IS	THE	TOILET	
	JUST	HERE		OA O Y
				OA O Y

Q4 r6

Q4 r9

Q4 r10

Q57 WHERE ARE YOU GOING VILLAGE O X Y O A Y
     TO THE NEXT O A X O Y

Q58 WHERE ARE YOU GOING STREET O X Y O A Y
     DOWN THAT O A X O Y

Q59 WHERE ARE YOU GOING KNOW O   O C
     I DONT O   O C

Q60 WHERE ARE YOU GOING O A O Y
     JUST HERE O A O Y

Q61 WHAT IS THE TIME O A X Y O A X Y
     QUARTER PAST FOUR O A X O Y

Q6/A8	WHAT	IS EARLY	THE NEXT	TIME WEEK	OCXY OCXY
Q6/A9	WHAT	IS	THE DONT	TIME KNOW	OAX C Y
Q7/A7	WHY	ARENT	YOU THOUGHT	DRESSED / WAS	O OC O
Q7/A9	WHY	ARENT	YOU DONT	DRESSED KNOW	O O

WHICH   IS DOWN   THE THAT   WAY STREET   O X YOC
                                              O X OC Y

WHICH   IS DOWN   THE THAT   WAY STREET   O X YOC
                                              O X OC Y

WHICH   IS I   THE DONT   WAY KNOW   O CC
                                              O OC

WHICH   IS JUST   THE HERE   WAY   OA O
                                              OA O

Q7
N7

Q8
N8

Q9
N9

Q10
N10

Q9 A2	WHERE	DO AT	THE	EAT TRANSPORCAFE	O X OA Y
					OAX O Y
Q9 A5	WHERE	DO AT	THE	EAT TERMINUS	O X OA Y
					OAX O Y
Q9 A6	WHERE	DO DOWN	THAT	EAT STREET	O X YOA Y
					OAX O Y
Q9 A9	WHERE	DO I	DUNT	EAT KNOW	O OC
					O OC
Q9 A10	WHERE	DO JUST	HERE	EAT	OA O Y
					OA O Y

Q10 A3	WHERE	CAN TO	I THE	GO NEXT	VILLAGE O X DAY Y OAX b Y
Q10 A6	WHERE	CAN DOWN	I THAT	GO STREET	O XY DA Y OAX b Y
Q10 A9	WHERE	CAN I	I DONT	GO KNOW	O OC O OC
Q10 A10	WHERE	CAN JUST	I HERE	GO	OA b Y OA b Y
Q14 A1	WHEN	DOES QUARTER PAST	IT	LEAVE FOUR	OAXYO OAX O Y
Q11 A8	WHEN	DOES EARLY	IT NEXT	LEAVE WEEK	OCXYO XY O X OC Y

Q11 A9	WHEN	OUFS 1	IT DONT	LEAVE KNOW

U |OC O

O |OC O

Q12 A1	WHEN	ARE QUARTER PAST	YOU	COMING FOUR

OAX Y|O

OAX |O Y

Q12 A6	WHEN	ARE EARLY	YOU NEXT	COMING WEEK

OCX Y|O XY

O X |OC Y

Q12 A9	WHEN	ARE 1	YOU DONT	COMING KNOW

O |OC O

O |OC O

DISCUSSION

GARVIN: I wanted to make a brief comment. I think that
what Margaret Masterman has been talking about raises a
fundamental question in the frame of reference of the history
of American linguistics, which is the following: Some years
ago, before generative grammars, there was an assumption
theory in American linguistics that intonations mark syntactic
boundaries. This is known as the famous intonational syntactic
marker.

What I understand from Margaret's discussion is that
intonations do not mark syntactic boundaries but they mark
what I would call lexical boundaries; that is to say, the
boundaries of major lexical units such as statements, if you
wish. You can call these lexical to differentiate them from
syntactic.

MASTERMAN: My view is there are two systems.

GARVIN: If this is a reasonable assumption, then this would
be, I think, worth pondering as a question of what it is that
intonation signals. And, of course, this raises a further
question in my mind; namely, whether there isn't a certain
confusion between intonation as a signal of lexical unithood
on the one hand, and semantic content of intonation on the
other. I think perhaps it's more the signal property of
intonation, and there is another difficulty here which is
that if you work with written texts you have to make some
very stong assumptions about consistency in reading in order
to use intonation as markers.

MASTERMAN: All this is quite true. I don't think it means,
however, that nothing can be done.

GARVIN: No. Lots can be done; on the contrary. I have
changed my mind from my original opinion that there is no
hypothesis to finding it very interesting. It is relatively
simple to detect the boundaries of short lexical units. You
can always decide what is a single lexeme by asking the ques-
tion, in pointing at an object, "What is this?", and the guy
will say "This is an ash tray", "This is a coffee cup," and
then you know that "ash tray" and "coffee cup," are single
lexemes. But if you want to know what are the larger units
in the lexicon, things that are more than single lexemes,
and how do you detect their boundaries, this has so far been
totally unanswered, and I think intonation may be one way of
marking lexical boundaries that linguists in this country
have overlooked at a time when they thought intonation was
important. At present, of course, the trend is so different
I don't know what most of us would consider significant.

HARPER: If we limit the discussion to rhythm texts, and to
the analysis of rhythm texts, I don't see at all what justi-
fication you have for saying that the phrases of these larger
units in written text are the same as those which evolve when
you read the text aloud. Do you have anything more to say
on that?

MASTERMAN: Of course the trouble is to get these things from
written text. Moreover, lots of study is needed to see what
different speakers do. I have been spending a lot of time
with different people reading aloud the same passage. They
are not as different as at first one feared. Pace is the main
difference. One man may put two phrases together, while
another has two separate ones.

GARVIN: That gives you another level of fusion.

MASTERMAN: Yes, it does, but it makes the hypothesis which I am sure is there less difficult.

VON GLASERSFELD: I would like to reinforce something Paul Garvin said which seems to have dropped under the table, and that is, the stress points in spoken text surely have some relation with the semantic content of the items that are being stressed. To see that, you only have to consider some artificially metered poetry like Latin poetry, which shows that very clearly. I don't know whether Margaret would agree with this, but I have the feeling that the study of the sounds and the stresses in spoken text is one way to leading toward delimitation of semantic, shall we call them, branches in a text. But the study of the content of certain items that coincide with the stress points by itself, without considering the stress, leads to the same delimitation.

MASTERMAN: Yes.

VON GLASERSFELD: I am not denying that the combination of both will be an extremely fertile one, but I believe some part of the goal can be achieved in another way.

MASTERMAN: I think this is quite right. Maybe we were rather stupid at the C.L.R.U., but we started by simply having linkage all alone, and then we found this wouldn't do. We needed a simplifying device. They heard me say we needed something to pick out all the stress words all the way down a piece of text. It's a game to see if the others can figure what is being said. If it is pronounced right you can.

V.

SOME QUANTITATIVE PROBLEMS IN SEMANTICS AND LEXICOLOGY

Stephen Ullmann
University of Leeds

In 1961 a symposium, rather similar to our own, was
held in Besançon on "The Mechanization of Lexicological
Researches". At that symposium, the Chairman, Professor
Quemada, distinguished between two groups of interests;
"classical lexicologists" who hoped to benefit by the new
machines for extending their possibilities of work conceived
along traditional lines, and "modern lexicologists", who, as
he put it, would never have entered the field without the
existence of new and powerful machines.

I come here quite unashamedly in the former capacity,
as a "classical lexicologist" who hopes that some of the
traditional and even perennial problems of semantics may be
solved, or at least more rigorously formulated, thanks to
computers and other aids, than has been possible so far.

Two points ought to be made quite plain from the very
outset. In this particular paper, the term "semantics" will
refer exclusively to lexical meaning; problems of meaning
arising below and above the word level will not be considered in
the discussion which follows.

The other point is this. In his well-known article on
"Computer Participation in Linguistic Research", (Language,
XVIII, 385-9) Paul Garvin distinguished between three degrees
of computer participation: "language data collection, which is
essentially a form of bookkeeping; computer programs using the
results of linguistic research; and automation of linguistic
research procedures." What I hope to talk about is at the
lowest level of this hierarchy. I do hope, however, to show

that while these problems seem very trivial from the computa-
tional point of view, their semantic and lexical implications
can be extremely useful and far-reaching.

I need hardly add a third point. I personally have
absolutely no expertise in computers, although I have had the
benefit of the advice of my colleagues in the University of
Leeds computing and data processing units.

A great deal has already been achieved in both descriptive
and historical semantics and lexicology with the aid of computers.
leaving aside such special applications as machine translation
and information retrieval, with the many semantic problems they
throw up, such as disambiguation, classification of concepts
etc. Computers have also been used, or could quite easily be
used, to tackle the crucial problem of all semantics, the
meaning of meaning, certain aspects of which may be quanti-
fiable. There are two factors in particular that are amenable
to such treatment: collocation and connotation. Collocation,
which looms large in the work of some British and American
linguists, is crying out for computer treatment. As regards
connotation, we have the famous Osgood experiment, with the
very misleading title The Measurement of Meaning, which is
really a measurement of connotation or emotive overtones. These
are very important applications with which I do not propose to
deal because they are already fairly well known. I should
rather like to consider another set of problems which seem
capable of being attacked with the aid of computers: certain
semantic and lexical phenomena which may be either synchronic
properties or diachronic tendencies. My basic assumption is
the existence of a research project like the one which
Professor Josselson and his team are engaged on, and it is
actually the privilege of talking to him and his fellow workers
last summer in Detroit which suggested to me the ideas which
follow. They are feeding two Russian dictionaries, published

at an interval of about twenty years from each other, into
a computer, and the main problem is to decide what to code.
There are eighty columns on a punched card, and only a small
portion will be used up by the immediate grammatical infor-
mation concerning each word. What else should we code from
the very outset, whether in a dictionary or in a corpus,
which is being fed into the computer? What would be the
semantic or lexical "parameters" that one might wish to code
and store in such a project?

What I have in mind is a code embodying as many semantic
and lexical criteria as possible. I am encouraged about the
feasibility of such coding by a recent book by S.H. Hollingdale
and G.C. Tootill on Electronic Computers (1965) where they
state that one of the desirable features of computer programming
techniques would be "the construction of programs so as to have
as wide a range of application as possible. The reason for
this is that a large program - which may take several months
to prepare and check thoroughly - represents a sizeable capital
investment and should be made to pay its way by being used to
the utmost." (p. 137)

I shall divide my suggestions into two groups; those
concerned with synchronic properties and those referring to
diachronic processes.

A. Synchronic Properties

As regards synchronic properties, for a long time
semanticists have been dealing with a variety of semantic
features whose relative frequency is characteristic of a
given language as opposed to other languages or as distinct
from earlier or later stages in its own development. Some
of these criteria, while very useful, are not precise enough
to be amenable to computer treatment, such as, for example,

the ratio of particular and generic terms. There are, howeve
others which could be quantified,but have not yet been quanti-
fied; linguists have so far relied on impressionistic hunches
and on a small number of examples.

There are four sets of synchronic problems which I should
like to discuss briefly: motivation, synonymy, polyvalency,
and semantic typology.

I. Motivation

The question of motivation, the contrast between conven-
tional and motivated, or opaque and transparent words, is a
perennial problem of linguistics and of the philosophy of
language, going back to classical antiquity, and reopened by
Saussure and more recently by Benveniste in the first number
of _Acta Linguistica_ (1939). There is a vast literature on the
subject which was recently surveyed in a useful bibliography
by Rudolph Engler in _Cahiers F. de Saussure_ (1962).

In this connection, one often hears impressionistic state-
ments of the kind: "German is more motivated than English or
French," and one is given some examples, very often the same
examples; where we say in German _Handschuh_, which is a motivated
compound, we say in English _glove_ and in French _gant_, which are
unmotivated, purely opaque and unanalysable terms. Where we
have in English _hippopotamus_, which is motivated only for
those who know Greek, in German one says _Nilpferd_, which is
intelligible to anyone who knows the name of the river _Nile_
and the German _Pferd_, "horse". There are also certain counter-
examples, like the one quoted by Uriel Weinreich in _Language_,
XXXI, p. 538. He points out that in the case of the English
grandson and the French _petit-fils_ we have motivated compounds,
whereas the corresponding German _Enkel_ is unanalysable. More-
over, Professor Weinreich rightly argues that, in view of the
quantitative nature of the problem, an uncontrolled list of
examples cannot serve as scientific evidence, and that it has

not been shown that the feature in question is necessarily characteristic of French.

It might be possible, by coding a synchronic dictionary or a corpus in the way I have suggested, to determine the ratio of motivated and unmotivated terms. Moreover, motivation is not a homogeneous phenomenon, and it would be interesting to know the relative frequency of its various types and subtypes in that particular corpus or dictionary. There are three different kinds of motivation. First of all, there is phonetic motivation or onomatopoeia, which may again be either primary or secondary. In the former the meaning itself is an acoustic phenomenon which is imitated by the sounds, as for example splash. In secondary onomatopoeia, it is some non-acoustic phenomenon, for example a movement or action, or a physical or moral quality, which is portrayed by the sounds, as in words like snip, snap, sneak, snoop etc.

Secondly, there is morphological motivation which is found in compounds and derivatives, and the latter may be further subdivided by having separate codings for those formed with prefixes. suffixes, or, in languages like Turkish, infixes.

Lastly, there is semantic motivation which has two sub-classes: metaphor and metonymy. The various possibilities which arise under motivation may thus be summed up in the following diagram:

Words
- I. Unmotivated
- II. Motivated
  - (1) Phonetically
    - (a) Primary onomatopoeia
    - (b) Secondary onomatopoeia
  - (2) Morphologically
    - (a) Compounds
    - (b) Derivatives
      - (i) Prefixes
      - (ii) Suffixes
      - (iii) Infixes
  - (3) Semantically
    - (a) Metaphor
    - (b) Metonymy
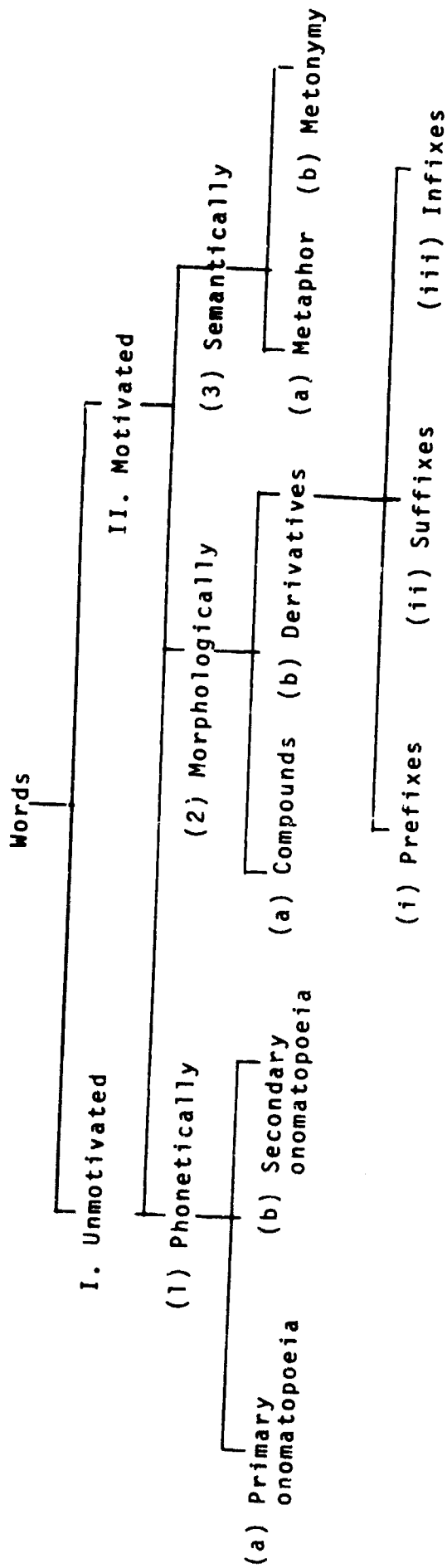
This diagram does not of course claim to be exhaustive or universally valid, and further distinctions will have to be made when dealing with particular languages.

Admittedly there are difficulties in this field. Before
one does the coding one will have to make certain decisions or,
to borrow a term from computer language, one may have to carry
out certain sub-routines. One may, for example, have to conduct
some psychological experiments, such as Wissemann and Chastaing
have devised in the field of onomatopoeia. In the matter of
morphological motivation, one will have to distinguish between
motivation within or outside the language. Thus hippopotamus
is not motivated from the English point of view: its motivation
lies in Greek; it is compounded of hippos "horse" and potamos
"river". Such formations may either have to be coded separately,
or they may, from the internal English point of view, be relegated
to Category No. I, that of unmotivated terms.

Motivation has some important educational implications.
One will teach a motivated language differently, establish
different associative relationships, than in teaching a less
motivated idiom. Even within one community, the use of learned
Graeco-Latin terms instead of transparent native formations may
erect what has been called a "language bar" between people with
and without a classical education. To the linguist, motivation
and its subclasses may also furnish valuable criteria for semantic
typology.

II. Synonymy

There are two aspects of synonymy which seem to be quanti-
tative in nature, but it is very difficult to see how the computer
could help in studying them. First there is the problem of
synonymic patterns: the organisation of synonyms into "double",
"triple" etc. scales. English has a double scale, "Saxon" versus
"Latin", in many cases: deep - profound, hearty - cordial; some-
times there is a triple scale: English, French, and Greek or
Latin: kingly, royal, regal. One wonders how frequent these
patterns are, but it is not easy to see how they could be coded
in a corpus or dictionary stored in a computer.

The other statistical concept in the sphere of synonymy is the concentration of synonyms in certain areas which bulk large in the interests of a certain community. For example, Jespersen counted in Beowulf thirty-seven different nouns for "hero" or "prince". If the computer could somehow help to identify these synonymic clusters, possibly by a system of cross-references, that would be great value to semantics, but one cannot immediately see how such phenomena could be coded on punched cards in the way motivation or polyvalency could be.

III. Polyvalency

In the field of polyvalency, where one and the same linguistic form has several different meanings, the crucial problem, which dictionaries often solve in a very inconsistent, arbitrary and haphazard way, is the distinction between homonymy and polysemy. In the case of polysemy, we have a single word with several senses. A classic example is the word operation which may be surgical, military, financial etc., according to the context. In the case of homonymy we have two, or more than two, forms which are identical but have different meanings and constitute different words, whether they belong to the same word-class or not, as for instance bear, noun, bear, verb, and bare. But there are a number of borderline cases, and all kinds of attempts have been made to find some precise formal criteria to separate homonymy from polysemy. The criteria which have been suggested include: rhyme; repetition; morphological and syntactic differences; the fact that a word may belong to more than one derivational series. But there are numerous cases where none of these criteria will help. Ultimately it is often a matter of the subjective criterion of Sprachgefühl, in so far as it can be reduced to some sort of precise control and measurement. Once again

Professor Weinreich has made a helpful suggestion: "Social
science",he points out, "has workable techniques for studying
opinions which could be applied to homonymy problems (if it
is granted that they are a matter of speakers' opinions) as
well as to political issues" (loc. cit., pp. 541-2). This
would involve a special subroutine for which the doubtful
cases would have to be identified by special coding.

Once one has been able to isolate the borderline cases,
the following alternatives would have to be coded separately.
We would have two basic types: unambiguous words and ambiguous
ones, the latter subdivided into homonymy and polysemy.
Homonymy can be further subdivided into three types whose
relative frequency in a given language it would be very
interesting to know: homophones, pronounced alike, but written
differently (bear - bare); homographs, like tear and tear,
spelled alike, but pronounced differently; finally, homonyms
stricto sensu, pronounced alike, written alike: page, boy,
and page of a book. In polysemy it would be quite possible
to enter against the word in the coding the number of meanings
in which it is used. In this way one could immediately test
the Zipf theory which claims that there is a correlation
between polysemy and word frequency. According to Zipf,
different meanings of a word will tend to be equal to the
square root of its relative frequency, with the possible
exception of a few dozen most frequent words, One would like
to have this widely tested, to see if there is anything like
a linguistic universal in this area.

The various alternatives which would have to be coded in
the field of polyvalency could be summed up as follows:

Words

I Unambiguous

II. Ambiguous

(1) Polysemy

(2) Borderline cases

(3) Homonymy

(a) 2 meanings

(b) 3 meanings

(c) n meanings

(a) Homophones

(b) Homographs

(c) Homonyms stricto sensu
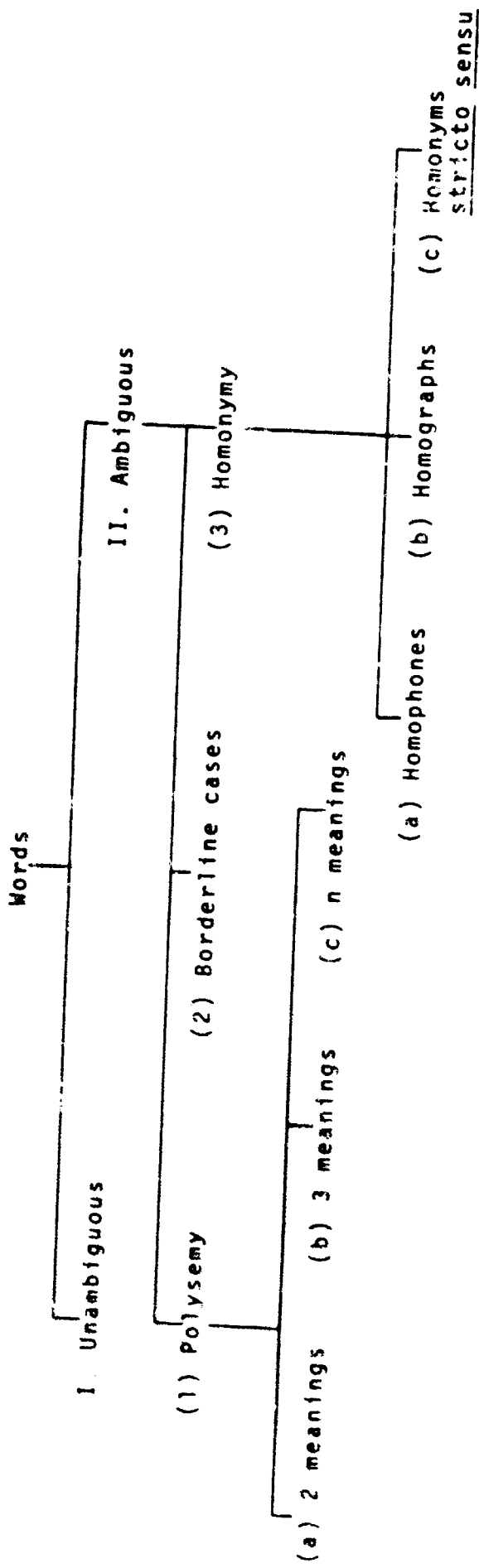
The connection between homonymy and word structure raises interesting problems. In Appendix I, I have reproduced some data from B. Trnka's now rather old book (1935), A Phonological Analysis of Present-Day Standard English. Needless to say, one would have to re-examine this material in the light of current phonemic analysis, but I doubt whether any significant differences would emerge. The table reveals some rather curious and unexpected correlations, which one would like to set against data in other languages. Thus, if one looks at the first of the fourteen types of monosyllables, those consisting of a single vowel ("a" in this system stands for a vowel and "b" for a consonant), one notices that there are only ten English words in this category, including five homonyms, whereas in French I have seen the figure of 52 mentioned.

IV. Typology

Motivation, synonymy and polyvalency are all potential criteria for semantic typology. Once precise figures are available for the relative frequency of each feature in various languages, the computer could test them for any possible correlations. Is there, as Bally has suggested (Linguistique générale et linguistique française, 3rd edition, 1950, p. 343), some kind of equilibrium between morphological and semantic motivation? Is there any connection between morphological motivation and polysemy? At present, one has certain hunches or subjective impressions which precise calculations might substantiate, correct or invalidate.

B. Diachronic Tendencies

A similar kind of coding, when applied to an historical or etymological dictionary stored in a computer, might throw light on the mechanism of semantic and lexical change. I

shall simply enumerate some features which seem to me capable
of this kind of treatment. Firstly, the relative frequency
of extensions and restriction of meaning. These are time-
honoured, traditional categories of semantic change, and
many scholars have suggested that restriction is more common
than extension. Is this true? An it it is true, does the
ratio very significantly from one language to another, or
from one period to another in the history of the same
language?

The various types of metaphor may also raise similar
problems. Are anthropomorphic metaphors, where we take parts
of our body and project them into the inanimate world around
us, more frequent than those working the other way around?
Are there any differences in this respect between various
languages and civilisations?

Synaesthetic metaphors, which are illustrated in Appendix
II, could also be quantified. These are transpositions of
sensations where two different sense-data are combined, as in
"sharp noise" where an adjective belonging to the sphere of
touch is used to characterise an acoustic experience. I have
done a certain amount of research on synaesthesia by examining
the usage of a dozen poets - French, English, American and
Hungarian - and my tentative findings have been corroborated
by subsequent studies in Italian and Rumanian, and I gather
from private correspondence, in Punjabi, Urdu and Persian.
On the graph in Appendix II, the figu⌐s to the right of the
slanting line refer to transfers from the lower senses to the
higher ones, whereas those to the left of it refer to "down-
ward" transfers from, say, sight to sound, or from sound to
touch. In nealy all the writers investigated, the "upward"
transfers were predominant; in over 2,000 examples there were
350 downward ones as against 1,650 upward. Touch was almost
invariably the commonest of sources, and sound the commonest
of recipients.

This kind of approach may also help us in linguistic reconstruction. Bloomfield has suggested that the traditional study of semantic changes "gives us some measure of probability by which we can judge of etymologic comparisons". That is to say, it shows how common or uncommon a change which we are inclined to posit may be. It may even enable us to choose between two alternative explanations. If we do not know which of two meanings came first, one relating to sound and the other to touch, then there is a strong probability, in accordance with the laws of synaesthesia I just mentioned, that the change occurred from touch to sound and not the other way round: a "sharp noise" is much more common and much more natural than a "noisy sharpness."

Changes in vocabulary may also be quantifiable, as shown in Appendices III and IV. Thus, the influx of French words into English was examined by Jespersen many decades ago. He took the first hundred words of French origin in the first nine letters of the Oxford Dictionary and the first fifty words of French original under J and L, and obtained some interesting results; his data showed, for example, a considerable "bulge" in the period 1250 to 1400, rather later than one might have surmised.

The intake of new words and meanings into English, studied in Thorndike's article on "Semantic Changes", is equally revealing. It is worth noting, for instance, that the two processes run, broadly speaking, along parallel lines; both show a peak period of productivity, from 1580 to 1620, and a "trough" from 1740 to 1780, followed by a certain revival of both creative processes.

One final example of lexical change: four French linguists, J. Dubois, L. Guilbert, H. Mitterand and J. Pignon, published in Le Français Moderne, 1960, a very interesting comparison of two successive editions of the Petit Larousse, that of 1948 and that of 1960, and they found quite considerable changes.

In 1948 there were 36,000 words in the dictionary. By 1960
over 5,000 had been omitted and nearly 4,000 had been added,
and there were so many additions or omissions of meanings
that about a quarter of the whole dictionary material had
changed. More than 100 new words from English had crept into
the language in the intervening twelve years. All these
problems could be profitably tackled with the help of computers.

It is clear from this brief survey, from which stylistic
problems have been deliberately excluded, that in semantics
and lexicology, the use of the computer enables us to tackle
old problems in a new way and, if one might say so, in a
more Cartesian way, "making everywhere such complete counts
and such general surveys that we should be certain not to
have omitted anything". These old problems might in their
turn throw up fresh ones, and the new approach may lead to
a much more precise formulation of semantic features and
tendencies than has been possible so far. This would help to
dispel any lingering doubts about semantics, which in the
immediate post-Bloomfieldian period were so widespread,
especially on this side of the Atlantic. Fortunately, there
has been a dramatic change during the last few years, thanks
partly to the emphasis on semantics in transformation theory
and generative grammar, but also thanks to many other new
initiatives of which the present symposium is a notable
example.

It has been suggested that semantics has at last begun
to come of age; if this is so, perhaps it is not fanciful to
hope that the computer may play a significant part in the
process.

Appendix I

WORD-STRUCTURE AND HOMONYMY IN ENGLISH

(from B. Trnka: A Phonological Analysis of Present-Day
Standard English)

TYPE	NUMBER OF PHONEMES	NUMBER OF WORDS	IN PER CENT	NUMBER OF HOMONYMS
1. a	1	10	0.31	5
2. ab	2	67	2.05	9
3. ba	2	174	5.37	91
4. bab	3	1,343	42.00	333
5. abb	3	28	0.87	2
6. bba	3	124	3.88	36
7. babb	4	433	13.45	53
8. bbab	4	709	22.46	105
9. bbba	4	19	0.59	1
10. babbb	5	14	0.43	-
11. bbabb	5	168	5.28	9
12. bbbab	5	75	2.36	5
13. bbabbb	6	3	0.09	-
14. bbbabb	6	11	0.34	-
1-14	1-6	3,178	100%	649

Appendix II

SYNAESTHETIC METAPHORS IN KEATS

	Touch	Heat	Taste	Scent	Sound	Sight	Total
Touch	-	1	-	2	39	14	56
Heat	2	-	-	1	5	11	19
Taste	1	1	-	1	17	16	36
Scent	2	-	1	-	2	5	10
Sound	-	-	-	-	-	12	12
Sight	6	2	1	-	31	-	40
Total	11	4	2	4	94	58	173

Appendix III

THE INFLUX OF FRENCH WORDS INTO ENGLISH

(from O. Jespersen: Growth and Structure of the English Language)

Before 1050	2	1451-1500	76
1051-1100	2	1501-1550	84
1101-1150	1	1551-1600	91
1151-1200	15	1601-1650	69
1201-1250	64	1651-1700	34
1251-1300	127	1701-1750	24
1301-1350	120	1751-1800	16
1351-1400	180	1801-1850	23
1401-1450	70	1851-1900	2
	581		1000

Appendix IV

INTAKE OF LIVE NEW WORDS AND MEANINGS INTO ENGLISH

(from E.L. Thorndike, "Semantic Changes", The American Journal
of Psychology, 1x, 1947, 588-97)

	WORDS	MEANINGS
OE	55	24
ME-1459	188	134
1460-1499	26	25
1500-1539	50	50
1540-1579	75	71
1580-1619	120	135
1620-1659	79	93
1660-1699	57	70
1700-1739	42	62
1740-1779	39	54
1780-1819	66	72
1820-1859	123	117
1860-1899	81	93
Size of Sample	9422	4101

DISCUSSION

WEINREICH: I wanted to ask you whether you had thought of another criticism which I had of your book on French semantics and which in a way is a criticism of traditional semantics at large, and should perhaps be reconsidered here. Suppose we wanted detailed quantitative data on the amount of motivation in a language. Every complex expression is motivated, so that if we are going to count it as motivated we have to have some simplex elsewhere in the language, or in another language, against which to select it.

For example, we will count "tablecloth" as motivated only because we need some criteria. Perhaps there are words in other languages with which we are very familiar which are simplex and therefore by contrast "tablecloth" is complex and would count as motivated. But what about "meeting room"? Is it an entity that we have to take into our calculations at all, or not? Can you suggest any criteria for something like this?

ULLMANN: That raises the whole issue or question of a sort of habitual colloquation. A set phrase becomes a compound. There are certainly criteria one can suggest for borderline cases: the actual intonational contour, sometimes a strong semantic shape, sometimes grammatical criteria. There is the famous example of "blackbird." Not all black birds are blackbirds. Sometimes there are grammatical criteria: Not " "I broke fast" but "I breakfasted this morning." But these don't usually help.

What I was trying to do was to code existing distinctions, not to write any. In analyzing a corpus one would have to make up one's mind. But I feel that at this very early and tentative stage, even taking a carefully prepared major dictionary such as, for example, the shorter Oxford Dictionary --

in other words as these French people have done -- and even
taking the question of arbitrariness which went into the
problem, which you rightly point out, it would still yield,
as to the law of large figures, quite interesting information,
at least interesting to me and to many other linguists.

VI.

PROBLEMS IN AUTOMATIC WORD DISAMBIGUATION

Herbert Rubenstein
Center for Cognitive Studies
Harvard University

Last year I explored the possibilities of automatic word disambiguation with the help of Janet Foster of Arthur D. Little, Inc.[1] This paper represents my recent thinking about the problems and results of our exploration.

Semantics is in its infancy, if not chronologically then certainly with regard to the paucity of its substance. We cannot hope for a useful comprehensive theory of semantics before the field has been limned out by some systematic accumulation of data. Before we go data gathering, however, it is essential to set some goals, attainable and delineated sharply enough so that we know when they have been achieved. I believe that automatic word disambiguation is a limited goal of this sort. I am using the word disambiguation to mean 'reduction of ambiguity' rather than 'total elimination of ambiguity.'

I take the research task to be this: to discover the the information necessary to enable a computer to take an isolated English sentence containing one or more homographs[2] and list all the meanings of the homographs acceptable to a native speaker and only those meanings. Note that the computer is not required to come up with a unique meaning unless, of course, the native speaker would accept only one meaning. Here is an example of a sentence containing four homographs:

The wire		upset		the wooden		coach
1. metal thread		3. overturned	5. made of wood		7. vehicle	
2. telegraph message		4. excited	6. awkward		8. trainer	

Since each of the homographs has two meanings, there are 16
possible combinations of meanings. We would want the computer
to indicate the three or four acceptable combinations: 1, 3,
5, 7; 1, 3, 6, 8; 2, 4, 6, 8; and possibly 1, 4, 6, 8. There
may be some reservation about the acceptability or about the
likelihood that a computer could recognize the last of these
since it involves an ellipsis: (The sight of) the metal thread
excited the awkward trainer.

Obviously such a computer program presupposes automatic
syntactic analysis. I am inclined to agree with Katz and his
collaborators (1963, 1964a, 1964b, 1965) that this must be a
transformational analysis since semantic rules can be success-
fully applied to the underlying kernels of a sentence.[3] Consider,
for example, the sentence The woman was fair in her treatment of
the workers. Obviously, if fair were analyzed as syntactically
associated with woman, fair would be incorrectly assigned 'light
in complextion' or 'pretty' as possible meanings in addition
to 'just.' Only in an analysis in which fair was associated
with treatment would it be properly interpreted only as 'just.'
The kernelization The woman treated the workers fairly would
be ideal for semantic analysis.

By presupposing a syntactic analysis program of this kind
we are able to bypass consideration of syntactic ambiguities
both in the surface structure, e.g., They (are flying) planes
versus They are (flying planes), as well as in the deep
structure, e.g., John is fit to teach, that is, 'John is fit
to be taught' and 'John is fit to teach others.'

There are obvious limitations on the kinds of language
that we could expect a computer to handle. Not only the
metaphoric languages of poetry but the make-believe of story

books and cartoons lies far beyond any reasonable expectation
for automatic word disambiguation. In the sentence The boxer
spoke well, we would expect boxer to be interpreted as 'pugilist'
not as 'kind of dog' despite the eloquence of Barnaby's Gorgon.
A frequent kind of ellipsis also beyond automatic semantic
analysis is that involved in the meaning 'representation of ----'.
For example, we would not expect the computer to interpret
plane as 'aircraft' in the sentence He put the plane in his
pocket, and yet the meaning 'toy aircraft' would be completely
acceptable to the human listener in many circumstances.

While there certainly are great difficulties involved in
developing a program for automatic word disambiguation, it is
worth noting that they are far less formidable than the diffi-
culties involved in realizing the goals set by Katz and his
collaborators: 1) to detect whether a sentence is uniquely
meaningful, ambiguous or anomalous; 2) to decide whether two
sentences are synonymous; 3) to decide whether a sentence is
analytic, synthetic or contradictory. Goals 2 and 3 require
complete semantic decomposition of all words and rules for
combining the elements of these decompositions. Goal 2 further
requires that a particular meaning is represented as composed
of the same elements regardless of the words used to express
that meaning. Word disambiguation, of course, does not require
such extensive semantic analysis but only the isolation of
those semantic elements which are useful in characterizing
the permissible environments of the various meanings of a
homograph. All in all, I believe that automatic word dis-
ambiguation is the simplest test of the feasibility of the
notion that lexical meaning can be at least partially analyzed
into components (semantic markers and selection restrictions
in Katz's parlance).

A word disambiguation program requires 1) a dictionary
in which each meaning of a word is listed together with all

the syntactic and semantic information pertinent to its
distribution; 2) rules governing the application of this
information. Stating the distribution of a meaning is
obviously a very difficult matter. Clearly the distribution
of a meaning of a word cannot be formulated in purely syntactic
terms but must be ultimately described in terms of the meanings
of the words with which it occurs. A great economy of des-
cription can be gained if this set of meanings can be character-
ized by a limited set of semantic elements. I shall use two
terms in speaking about semantic elements which are related
much as _phoneme_ and _phone_ are to each other: _semantic components_
are those elements of meaning whose utility for word disambigu-
ation has been established according to various criteria;
semantic features are elements whose utility remains to be
established,

The making of this dictionary may be facilitated by two
fairly reasonable assumptions: first, the meaning of a homo-
graph depends upon the meaning of a word which stands in one
of a limited set of syntactic relations to the homograph. Such
relations are: noun-pronoun, adjective-noun, noun-noun, adverb-
verb. There are also several tripartite relations, e.g.,
subject-verb-object, noun-preposition-noun, verb-preposition-
noun. Disambiguation is not accomplished within relations
like preposition-adjective, adverb-noun or subject-object.
The second assumption is that in most instances it is a noun
meaning that selects the meaning of a homograph. Thus in
general there is no need to decompose the meanings of non-nouns
but merely to state the semantic components of the nouns with
which these non-noun meanings occur. There are some instances,
however, where semantic components of verbs play a role. The
disambiguation of the prepositions, for example, requires infor-
mation about the meanings of the noun and verb with which it
is used. Disambiguation of adverb homographs (which are few

since they often become monosemous in derivation from a homo-
graphic adjective) also may require information about the
meaning of the co-occurring verb. The implications of these
assumptions then for the structure of the dictionary are the
following: Adjective and verb meanings would be followed by
the components of the meanings of the nouns with which they
may occur. For the transitive verb there would have to be
information both about the subject and object noun meanings.
Adverb meanings would be followed by the semantic components
of the verbs with which they may occur. Most preposition
meanings will probably require the components of both co-
occurring nouns and verbs. Only for noun entries would the
meaning be followed by semantic components derived from its
own features.

The dictionary would also include non-homographs since
their characterizations can serve to disambiguate co-occurring
homographs as we shall see in examples below.

The rule for the application of dictionary information
would be of this general form: meanings of words within
syntactic relations like those cited above are compatible
unless they are contradictory on any semantic component. By
contradictory I mean that the meaning of one word has (+) on
an element where the meaning of the other word has (-). Note
that ($\pm$) is compatible with (+), (-) or ($\pm$).

Examples of disambiguation in various syntactic relation-
ships:

The illustrations are, of course, incomplete. Not
all meanings are given nor is any meaning completely characterized.
The components, which are shown in brackets, are only tentative.
Note that components following non-nouns are descriptive of the
meanings of the nouns with which they occur and not of the
meanings of the non-nouns.

In strings of components, the comma = intersection,
and or = exclusive or.

(1)   N Adj

(1a) The shawl was blue.

shawl $[\overline{+}$ physical obj.$]$, $[\overline{-}$ person$]$

blue$_1$ 'color' $[\overline{+}$physical obj.$]$, $[\pm$ person$]$

blue$_2$ 'melancholy' $[\overline{+}$ physical obj.$]$, $[\overline{+}$ person$]$ or
   $[\overline{-}$ physical obj.$]$, $[\overline{+}$ intellectual product$]$

Acceptable: shawl blue$_1$

(1b) The bark was soft.

bark$_1$ 'animal sound' $[\overline{-}$ physical obj.$]$, $[\overline{+}$ sound$]$

bark$_2$ 'cortex of plant' $[\overline{+}$ physical obj.$]$

soft$_1$ 'not loud' $[\overline{-}$ physical obj.$]$, $[\overline{+}$ sound$]$

soft$_2$ 'not hard' $[\overline{+}$ physical obj.$]$

soft$_3$ 'not difficult' $[\overline{-}$ physical obj.$]$, $[\overline{-}$ sound$]$

Acceptable: bark$_1$ soft$_1$, bark$_2$ soft$_2$

(2)   NV

The sap is running.

sap$_1$ 'plant juice' $[\overline{+}$ natural liquid$]$

sap$_2$ 'fool' $[\overline{+}$ animate$]$, $[\overline{+}$ person$]$, $[\overline{+}$ having legs$]$

run$_1$ 'move rapidly on legs' $[\overline{+}$animate$]$, $[\pm$ person$]$,
   $[\overline{+}$ having legs$]$

run$_2$ 'flow' $[\overline{+}$ natural liquid$]$

Acceptable: sap$_1$ run$_2$, sap$_2$ run$_1$

(3) NVN

The boxer passed the ace.

boxer₁ 'pugilist' /⁺ animate7, /⁺ person7 /⁺ having hands7
                   /⁺ mobile7

boxer₂ 'kind of dog' /⁺ animate7, /⁻ person7 /⁻ having hands7
                   /⁺ mobile7

pass₁ 'hand'   Subj. /⁺ having hands7; Obj. /⁺ physical
                   object7, /⁺ portable7

pass₂ 'go by'   Subj. /⁺ mobile7; Obj. /⁺ physical obj.7

pass₃ 'give satisfacory grade to'
                   Subj. /⁺ person7; Obj. /⁺ physical obj.7

ace₁ 'highly proficient person'
                   /⁺ animate7, /⁺ person7, /⁺ physical obj.7,
                   /⁻ portable7

ace₂ 'playing card' /⁻ animate7, /⁻ person7, /⁺ physical obj.7,
                   /⁺ portable7

Acceptable: boxer₁ pass₁ ace₂; boxer₁ pass₂ ace₁;

            boxer₁ pass₂ ace₂; boxer₁ pass₃ ace₁;

            boxer₁ pass₃ ace₂; boxer₂ pass₂ ace₁;

            boxer₂ pass₂ ace₂.

The partial dependence of the meaning of <u>pass</u> on subject and object is shown by the fact that <u>pass</u>₁ can be the interpretation only if the subject is marked /⁺ having hands7 and the object is marked /⁺ portable7.

(4) V Adv.

(4a) He grasped it roughly.

<u>grasp</u>₁ 'seize' Obj. /⁺ physical obj.7; V /⁺ contact7
                     (derived from verb meaning)

<u>grasp</u>₂ 'understand' Obj. /⁺ physical obj.7; V /⁻ contact7

<u>roughly</u>₁ 'not delicately' /⁺ physical obj.7, /⁺ contact7

<u>roughly</u>₂ 'incompletely' /⁺ physical obj.7, /⁻ contact7

Acceptable: grasp₁ roughly₁, grasp₂ roughly₂

(4b) He spoke sharply

    __speak__ 'utter' $\quad$ [∓ communicate] (derived from verb

    __sharply$_1$__ 'angrily' $\quad$ [∓ communicate] meaning)

    __sharply$_2$__ 'in fashion' [⁻ communicate]

    Acceptable: speak sharply$_1$

    Compare: he dressed sharply$_2$

(5) N_1 __be__ Prep N_2

(5a) There was a lecture about the room.

    __about$_1$__ 'concerning' N_1 [∓ communication]; N_2 unspecified

(5b) There was dust about the room.

    __about$_2$__ 'around' N_1 [⁻communication/; N_2 [∓ location]

    The distance was __about__ a mile.

    __about$_3$__ 'approximately' N_1 unspecified; N_2 [∓ quantity]

(6) V Prep N_2

(6a) He drove by the hospital.

    __by$_1$__ 'past' V [∓ locomotion] (derived from verb meaning);

                N_2 [∓ physical obj.]

(6b) He worked by the hospital.

    __by$_2$__ 'near' V [⁻ locomotion]; N_2 [∓ location]

(6c) He worked (drove) by the rules.

    __by$_3$__ 'according to' V unspecified; N_2 [⁻ physical object]

Examples of disambiguation of noun homographs by monosemous non-nouns

    NV The barbet scared $\qquad$ [∓ able to fly].

    VN The man frequented the bar $\qquad$ [∓ location].

    N Adj. His bishop was foamy $\qquad$ [∓ potable].

The task of finding semantic components bears at least
a superficial resemblance to the task of finding distinctive
features in phonology. Both involve the problem of segmenta-
tion: How to divide the stream of speech? Shall we consider
a piece of meaning as one or two potential semantic components
/natural liquid/ or /natural/ and /liquid/? And both present
the difficulty of discovering the commonality in the members
of a set which has been defined distributionally. The dif-
ferences between the tasks, however are more impressive than
the similarities. First, there is a large quantitative dif-
ference. The number of distinctive features in a language
lies between 8 and 12. The number of semantic components
required for word disambiguation may come to 100 or more. This
relatively small number of distinctive features together with
the accumulated knowledge of the phonologies of a large number
of languages serves to simplify the linguist's task of describing
the distinctive features of a previously uninvestigated language.
He relies on his experience and assumes, at least tentatively,
that an acoustic phenomenon is not a distinctive feature unless
it is known to have served in this role in some other language.
In semantics I believe that some components may turn out to be
unique to particular languages since homophony which produces
a substantial portion of the words with more than one meaning,
is the result of phonological change and is, in general,
unaffected by the meanings of the morphemes involved. Secondly,
distinctive features and semantic components differ in the
nature of their referents. A distinctive feature refers to a
class of physical events which are part of the natural speech
process. A semantic component is an expression of some part
of a set of meanings, which, even in their most physical form,
are utterances about language. This implies, since there is
little constraint on the form of such utterances, that the
reliability with regard to the way in which a meaning is

explicated would be quite low despite the fact that there must be a high degree of agreement within any linguistic community on what words or sentences mean. Thus one of the main problems in semantic analysis is the development of procedures which allow the investigator to disregard formal differences in statements of meaning or parts of meaning. A very trivial example -- we would not want to attach any importance to the fact that a semantic feature of boy was labeled /person/ or that it was labeled /human/.

We now come to the crux of our problem. How do we obtain the semantic components, i.e., those semantic features that serve to disambiguate homographs?

Our experience suggests the following procedure:

1. Consider a kernel sentence in which the meaning of a noun disambiguates a homograph. The noun does not have to be monosemous since its meaning in the sentence is given.

We started with the simplest types of kernels and proceeded to the more complex, i.e., we considered types like $N V_i$, N be Adj; then went on to types like $N V_t N$, $N V_i$ Prep N, $N V_t$ N Prep N. Example: (1) The man is running. run_1 'move rapidly on legs'

2. Substitute other nouns of a wide variety of meanings for the disambiguating noun in the given kernel. These substitutes must select the same meaning of the homograph as the original noun.

Example:

The following are some of the possible substitutes for man together with some of their semantic features:

(1a) man /animate/, /natural/, /person/, /male/, /adult/, /having two legs/

(1b) girl /animate/, /natural/, /person/, /female/, /having two legs/

(1c) mouse /animate/, /natural/, /animal/, /mammal/, /having four legs/

(1d) lizard /animate7, /natural17, /animal7, /reptile7, /having
four legs7

(1e) beetle /animate7, /natural17, /animal7, /insect7, /having six
legs7

(1f) spider /animate7, /natural17, /animal7, /arachnid7, /having
eight legs7

(1g) sandpiper /animate7, /natural17, /animal7, /bird7, /having
two legs7

Another interpretation of sentences like The man is running
may come to mind; namely, 'the man is a candidate.' I have
excluded this interpretation because it seems to me to belong
to the province of elliptical language, that is, this interpre-
tation comes to mind only if one completes the sentence with
some phrase like for office.

3. Consider kernels of the same syntactic form as the original
in which the homograph has different meanings. Obtain a kernel
for each different meaning of the homograph. Go through Step 2
with each of these kernels.

Example:

(2) The smelt are running. run$_2$ 'migrate in large schools'

(2a, b, c) smelt, salmon, tuna, /natural17, /animate7, /fish7

(3) The sore is running. run$_3$ 'secrete fluid'

(3a) sore /inanimate7, /natural17, /body part7, /acquired7

(3b) eye /inanimate/, /natural17, /body part7, /visual organ7,
/congenital17

(3c) nose /inanimate7, /natural17, /body part7, /olefactory
organ7, /congenital17

(4) The water is running. run$_4$ 'flow'

(4a) water /inanimate7, /natural17, /liquid7, /HOH7, /for
drinking or washing7

(4b) sap /inanimate7, /natural17, /liquid7, /juice of plant7

(5) The ink is running. run_5 'spread'

   (5a) ink [inanimate], [liquid], [coloring matter], [for writing]

   (5b) dye [inanimate], [coloring matter], [for coloring material]

(6) The stocking is running. run_6 'unravel'

   (6a) stocking [inanimate], [artifact], [knitted of fine thread], [clothing for legs]

   (6b) lingerie [inanimate], [artifact], [knitted of fine thread], [underclothing]

(7) The motor is running. run_7 'operate in place'

   (7a) motor [inanimate], [artifact], [stationary], [having rotating part], [for imparting motion]

   (7b) fan [inanimate], [artifact], [stationary], [having rotating part], [for making breeze]

   (7c) refrigerator [inanimate], [artifact], [stationary], [having rotating part], [for cooling something]

(8) The streetcar is running. run_8 'go on schedule'

   (8a) streetcar [inanimate], [artifact], [vehicle], [scheduled], [public], [electric], [land], [surface], [on tracks]

   (8b) subway [inanimate], [artifact], [vehicle], [scheduled], [public], [electric], [land], [subsurface], [on tracks]

   (8c) bus [inanimate], [artifact], [vehicle], [scheduled], [public], [gasoline powered], [land], [surface]

   (8d) ferry [inanimate], [artifact], [vehicle], [scheduled], [public], [water], [surface]

4. icking all the kernels in which the homograph has the
same meaning as a set, consider what semantic features are
common to all the nouns in the set.

Example: Features common to the noun subjects of run:

   run$_1$ $\underline{/}$animat$\underline{e}\overline{/}$, $\underline{/}$natura$\underline{l}\overline{/}$, $\underline{/}$having leg$\underline{s}\overline{/}$
   run$_2$ $\underline{/}$animat$\underline{e}\overline{/}$, $\underline{/}$natura$\underline{l}\overline{/}$, $\underline{/}$fis$\underline{h}\overline{/}$
   run$_3$ $\underline{/}$inanimat$\underline{e}\overline{/}$, $\underline{/}$na'ura$\underline{l}\overline{/}$, $\underline{/}$b$\underline{c}$dy pa·$\underline{t}\overline{/}$
   run$_4$ $\underline{/}$inanimat$\underline{e}\overline{/}$, $\underline{/}$natura$\underline{l}\overline{/}$, $\underline{/}$liqui$\underline{d}\overline{/}$
   run$_5$ $\underline{/}$inanimat$\underline{e}\overline{/}$, $\underline{/}$coloring matte$\underline{r}\overline{/}$
   run$_6$ $\underline{/}$inanimat$\underline{e}\overline{/}$, $\underline{/}$artifac$\underline{t}\overline{/}$, $\underline{/}$knitted of fine thread$\underline{d}\overline{/}$
   run$_7$ $\underline{/}$inan.mat$\underline{e}\overline{/}$, $\underline{/}$artifac$\underline{t}\overline{/}$, $\underline{/}$stationar$\underline{y}\overline{/}$, $\underline{/}$having rotating par$\underline{t}\overline{/}$
   run$_8$ $\underline{/}$inanimat$\underline{e}\overline{/}$, $\underline{/}$artifac$\underline{t}\overline{/}$, $\underline{/}$vehicl$\underline{e}\overline{/}$, $\underline{/}$schedule$\underline{d}\overline{/}$, $\underline{/}$publi$\underline{c}\overline{/}$

5. Eliminate any common feature found in more than one set. If,
as a result of this restiction, it turns out that all the features
common to a set have been eliminated, there are several possible
courses of action: (a) Examine set for other possible common
features which may be unique to its members. (b) Reconsider the
segmentation of the common features. In our example if $\underline{/}$liqui$\underline{d}\overline{/}$
were a feature of dye, it would be a common feature of Set 5 as
well as of Set 4, and Set 4 would have no unique common feature.
A possible solution then would be to treat the features $\underline{/}$natura$\underline{l}\overline{/}$,
$\underline{/}$liqui$\underline{d}\overline{/}$ as a unit, $\underline{/}$natural liqui$\underline{d}\overline{/}$ which would serve as a common
feature unique to Set 4. (c) Reconsider whether the meaning of
the homograph selected by the set in question is truly distinct
from all the other meanings of the homograph. Indeed if there
is no environment that is unique to a particular meaning, it
is unlikely that we are dealing with a distinct meaning of the
homograph. This requirement that the set of selectors have at
least one common feature, unique to the set, provides us with
a check on our intuition regarding the distinctness of meanings.

Example: Tentative semantic components

   run$_1$ $\underline{/}$having leg$\underline{s}\overline{/}$
   run$_2$ $\underline{/}$fis$\underline{h}\overline{/}$
   run$_3$ $\underline{/}$body par$\underline{t}\overline{/}$
   run$_4$ $\underline{/}$natural liqui$\underline{d}\overline{/}$

run$_5$ /coloring matter_/
run$_6$ /clothing, knitted of fine thread_/
run$_7$ /stationary, device with rotating part_/
run$_8$ /vehicle, scheduled, public_/

I had considered making Step 5 more restrictive, that is, eliminating any feature which occurred in members of different sets even if it was common to all the members of only one set. The motivation for this lay in the view that components should bi-uniquely identify the members of a distribution class (all and only word-meanings[4] occurring in environment X, a meaning of a homograph, have /y/). This would be intuitively satisfying for the notion that /y/ selects X. However, this view cannot be maintained in face of the fact that the same word-meaning may occur with different meanings of the same homograph. Consider the sentence The men took the train. In the more frequent interpretation of this sentence, take$_1$ would have the meaning 'ride as passengers;' however, in another interpretation take$_2$ could mean 'take possession of.' The common features of noun objects of take$_2$ e.g., train, bus, ferry, /vehicle_/, /public_/, /scheduled_/ would consequently all be eliminated as tentative components since all these nouns as well as many others can occur as objects of take$_2$. Since the subject nouns of take in both meanings are the same, take$_1$ would have no components and consequently could not have a separate listing in our dictionary but would be included in some other meaning like 'take possession of,' 'carry,' etc. The psychological reality of the meaning 'ride as passenger' is clearly attested by the jarring effect of a sentence like We would have taken the train to Washington but it was too heavy. If, however, we adopt the weaker rule as presented in Step 5, our dictionary would show information like the following for objects of these two meanings of take:

take$_1$ 'ride as passenger' /∓ vehicle_/, etc.
take$_2$ 'take possession of' /± vehicle_/, etc.

The (+) indicates that the object to take$_2$ may or may not have /vehicle/ as part of its meaning. Obviously then kernels like The men took the train would still be recognized by our program as ambiguous, but this is what we wanted since such kernels would be ambiguous to humans if they were presented in isolation.

6. You will note that in listing the tentative semantic components above, I have bracketed them so that there is apparently only one component associated with each meaning of run. Since Occam's injunction Entia non sunt multiplicanda practer necessitatem hangs heavily above us, we shall assume that the features within brackets are not independent until we learn otherwise.

Discovering independence proceeds as one applies Steps 1-5 to other homograph sets. For example

(1)	He took a train.	take$_2$ 'ride as a passenger'
(1a)	train, plane, ferry	/+ vehicle, public, scheduled/
(1b)	taxi, rickshaw	/+ vehicle, public/, /- scheduled/
(2)	He took the car.	take$_2$ 'ride at the controls of a vehicle'
(2a)	car, rowboat	/+ vehicle/, /- public/, /- scheduled/

We may consider the possibilities in matrix form:

/vehicle/	/public/	/scheduled/
+	+	+
+	+	-
+	-	-

The matrix shows that all three features are at least partially independent and so we would at this point revise our dictionary entry, train, etc. to read /+ vehicle/, /+ public/, /+ scheduled/.

FOOTNOTES

1. My work at the Center for Cognitive Studies was carried on under ARPA Contract SD-187. Mrs. Janet Foster was supported by Contract AF 19 (628) - 3311 monitored by the Decision Sciences Laboratory, Electronic Systems Division, USAF.

2. The term <u>homograph</u> is used in this paper to refer to any word with more than one meaning whether this resulted from phonological change or not.

3. No distinction will be made here between kernel and basic string. See Chomsky (1965) especially pp. 17, 18.

4. By the expression <u>word-meaning</u> I mean a particular string of phonemes constituting a word together with its particular grammatical category and lexical meaning.

BIBLIOGRAPHY

Chomsky, N. Aspects of the Theory of Syntax. Cambridge,
    Massachusetts: MIT Press, 1965.

Katz, J.J., and J.A. Fodor. "The Structure of a Semantic
    Theory." Language, 39, pp. 170-210, (1965).

——————— . "Semantic Theory and the Meaning of 'Good'."
    J. of Philosophy, 61, pp. 739-766, (1964a).

——————— , and P. Postal. An Integrated Theory of
    Linguistic Descriptions. Cambridge, Massachusetts:
    MIT Press, 1964b.

——————— . Philosophy of Language. Mimeographed, 1965.

DISCUSSION

YNGVE: In speaking about resolution of ambiguity and elimination of senses, I would propose to restict the word "disambiguation" to the procedure envisaged in the Katz-Fodor theory. We all realize that the general idea of matching of components, as you call them, is far older than that, but let's use "disambiguation" for the exact scheme of Fodor-Katz.

Now, my question is, in that sense is it disambiguation? Are you following exactly the Fodor-Katz scheme?

RUBENSTEIN: It is precisely what I hoped to bypass. One of Katz's goals is the business of saying whether two sentences are synonymous. To do this, then, you would have to express them both in some fcrm, different from either of the sentences directly, and then match this to the meta-lingusitic expression of the content of the sentences.

To do that, it means that you have to have total decomposition, and I have not done that. I am using the senses that one might normally have in a dictionary.

My concern, so far as components go, is that these components match pretty closely the general notion of what he calls selection features or selection restictions.

BAR-HILLEL: I am sorry that Katz isn't here, because I would like to make the following very stong statement; namely, that the procedure proposed by Katz and Fodor is nothing but an adaptation of the first, of my own, in 1953. I take the responsibility for that part. But nevertheless, it might have been at that time a good try, but I think in 1965 it is not even a good try at all.

I started to say in my own presentation that I think this whole view of a sense being a bundle of sematic features

is utterly unacceptable except as a rough approximation on
a very limited sub-set of cases, but certainly not beyond
that.   So any attempt to impose this view on the totality
of semantics must wind up with total disaster.

GARVIN: Why?

BAR-HILLEL: As I tried to show, because the dictionary
entries could only cover a very small part of the so-called
meaning rules which go far beyond the semantic components or
semantic features.

VII.

SOME SEMANTIC RELATIONS IN NATURAL LANGUAGE

Ferenc Kiefer

Computing Centre of the
Hungarian Academy of Sciences

In this paper I want to show that - by defining various
semantic relations between words[1] - similarity is a basic
semantic relation because some of the other semantic relations
can be traced back to the former one. On the other hand, it
seems obvious that the semantic relations involve a hierarchical
structure of semantic categories, therefore the semantic
relations are defined in a way that such a system of semantic
categories is taken for granted. The claim that the semantic
relations between sentences depend on the semantic relations
between the words constituting the corresponding sentences can
be justified by using a well-defined conceptual apparatus.

1. Let us consider a set of categories K where the notion of
"category" is taken as a primitive notion. K must meet two
requirements.
  (i) it must be finite;
  (ii) the categories of K must be linguistically relevant in
      a certain way[2].
There seems to be no formal way of distinguishing "grammatical"
and "semantic" categories. Since there are well-known reasons
in support of a - not necessarily strict - distinction between
grammar and semantics we may proceeds by postulating two distinct
subsets of K, the set of grammatical categories K_G and the set
of semantic categories K_S, so that

$$K = K_G \cup K_S$$

(and of course, $K_G \sqcap K_S = 0$).[3]

These two sets are far from being unordered. On the contrary, the deeper we penetrate into the semantics of natural language the more structured the set of semantic categories seems to be. To put it differently, the more facts about language we want to describe by means of semantic categories, the more complicated structure we have to impose on K_S. (It seems to me that the structure of K_G will not be considerably increased this way, so the structure of K_G is much simpler than that of K_S.) So far we do not know how complicated the structure of K_S really is. The first thing we already know is that there are at least two basic relations characterizing both K_G and K_S and that there are some others (see below) that refer only to K_S.

1.1 There is undoubtedly a hierarchy between the categories. If we introduce an arbitrary category C_1^0 characterizing each element out of the vocabulary of a given language, then we have the following configuration:

(iii)
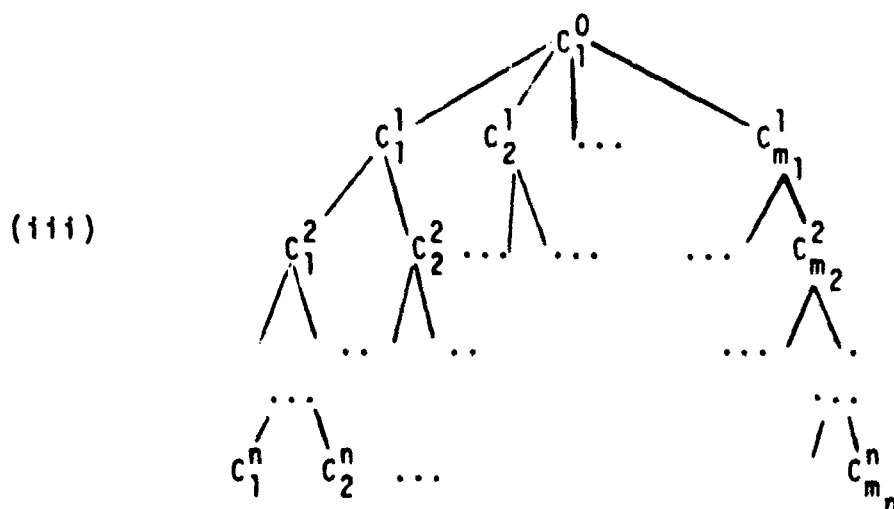
where n ($n \geq 0$) indicates the n-th level in the hierarchy
and m_i ($m_0 = 1$, $1 \leq m_i \leq m_n$) stands for the number of
categories on the i-th ($0 \leq i \leq n$) level[4].

It may be assumed that
$$1 < m_1 < m_2 < ... < m_n \quad .$$
in other words in general a category falls into one or more
subcategories on the next level.

So far we have not decided the q stion as to whether
there exists one such system as (iii) or several ones and,
on the other hand, whether both grammatical and semantic
categories are involved in a given system (iii). There
are good reasons for setting up hierarchical systems like
(iii) both for the elements of K_G and of K_S separately.
Although very little is known concretly as to the categories,
it seems evident that quite a few semantic categories would
occur more than once in a system including all categories[5].
On the other hand, it might be the case that we want to
compare words belonging to different grammatical categories[6].

In the following we shall not bother about grammatical
categories and assume that we have a hierarchical system like
(iii) of semantic categories at our disposal. Let us further
assume that it is possible to assign to words that may be
characterized semantically at least one category of each level
in (iii). More precisely, any word must be positively or
negatively specified with respect to at least one category
on each level. Henceforth such an assignment will be referred
to as the semantic characterization of the given word.

It is not quite clear to what a degree systems like (iii)
are universal. Any statement with respect to questions concerning
the universal character of (iii) cannot be seriously considered
at the present stage of our knowledge.

1.2 The other basic semantic relation between categories
may be referred to as inclusion. By <u>inclusion</u> we understand
the following relation. If C_i and C_j are two categories
and if w is a given word, further if whenever w is characteri-
zed by C_i, it is at the same time characterized by C_j as
well, then C_j includes C_i. We designate this relation by
" $\Longrightarrow$ ". According to the above

(iv) $\qquad C_i \Longrightarrow C_j$.

Generally we have a chain of "included" categories, i.e.

(v) $\qquad C_i \Longrightarrow C_{i+1} \Longrightarrow \quad \ldots \Longrightarrow C_{i+k}$ [7] .

One might think of imposing on (iii) thoroughout the
relation (iv), i.e. to require that

(vi) $\qquad C_i^p \Longrightarrow C_j^{p-1}$

for $1 \leq p \leq n$, $1 \leq i$, $j \leq m_k$ and $1 \leq k \leq n$. This requirement
could not be evidently met if we take only one system (iii)
for granted (i.e. including all grammatical categories as
well)[8]. However, if we take just one system (iii) of semantic
categories and leave aside grammatical categories, (vi) might
be put as a general requirement.

1.3 The semantic characterization of words may be visualized
as a labelled tree (which, of course, is not in general a
subconfiguration of (iii) having as many paths as the word under
consideration has meanings[9]. We will assume that words having
various "part-of-speech" categories are characterized in-
dependently, i.e. we assign to a given word as many different
labelled trees as to how many "part-of-speech" categories it
belongs. In the following, here, too, we leave grammatical
categories out of consideration.

2. Similarity[10]

2.1 Let n be the number of levels in (iii). Two words, x and y, are said to be similar on th j-th level and with respect to the i-th path, if and only if, their characterization on the i-th path coincides in the j-th category. Formally

$$x \underset{i,j}{\approx} y \quad .$$

Two words, x and y, are said to be fully similar on the j-th level if and only if the characterization of the two words containsthe same number of paths and if

$$x \underset{i,j}{\approx} y$$

for every $1 \leq i \leq r$, where r stands for the number of paths.

It should be noted that similarity is an equivalence relation[11].

Two words, x and y, are said to be first order similar on the i-th path, if and only if their corresponding characterizations coincide onthe first level. Formally

$$x \overset{1}{\underset{i}{\approx}} y \quad .$$

Two words, x and y, are said to be k-th order similar on the i-th path, if and only if

$$x \overset{k}{\underset{i}{\approx}} y$$

for every $1 \leq k \leq n$, where n stands for the number of levels in (iii).

If exactly k = n, then the k-th order similarity on the i-th path may be called <u>synonymy on the i-th path</u>.

Similar definitions - mutatis mutandi - are valid for the full similarity, i.e.

two words, x and y, are said to be first order similar, if and only if both x and y have in their characterizations the same number of paths and

$$x \overset{1}{\underset{i}{\approx}} y$$

for every $1 \leq i \leq r$, where r stands for the number of paths.

Two, words, x and y, are said to be k-th order similar, if both x and y have in their characterization the same number of paths and

$$x \overset{k}{\underset{i}{\approx}} y$$

for every $1 \geq i \leq r$ and $1 \leq k \leq n$, where r stands for the number of paths and n for the number of levels.

Is exactly $k = n$ and $i = r$, then the relation between x and y may be referred to as <u>full synonymy</u>.

2.2 By way of illustration let us consider a few examples.

The words "boy" and "man" are simialr on a certain j-th level and with respect to at least one path, because they have at least one category in common, let us say the category "Male". The two Hungarian equivalents for "dog": "eb" and "kutya" are fully similar on at least some of the paths and syonymous on at least one path (maybe fully synonymous). Two German equivalents ior "slippers": "Hausschuhe" and a South-German word "Schlappen" are probably fully synonymous.

On the other hand, words like "man" and "woman" differ already on a higher level, while "boy" and "man" will still

coincide on this level. To put it in our terminology, the order of similarity is lower in the case of "man" and "woman" than in the case of "boy" and "man".

2.3 It goes without saying that the similarity relation defined in the above way might be considered as a basis for comparision of sentences. As a first approximation we could restrict ourselves to so-called copula-type sentences or even more to transformationally not compound copula-type sentences like

(vii)   Peter is tall.

(viii)   Peter is clever.

(ix)   Peter is corpulent.

(x)   Peter is wise.

(xi)   Peter is skillful.

(xii)   Peter is dexterous.

Already a superficial inspection of sentences (vii) - (xii) reveals the fact that the sentences (xi) - (xii) are closer to each other than the sentences (vii) - (x) and that there is a similarity in the above sense between (vii) - (ix) and (viii) (x), the latter being even similar to (xi) - (xii) in a way. I believe that this similarity can be only formulated in terms of categories. Of course, I am quite aware of the difficulties that arise by comparing sentences. Firstly, there are a great number of sentences (in fact, infinitely many) which are not similar in any way. However, if two sentences reveal similarity, then this should be formulated in exact terms. Secondly, the comparison will become extremely complicated in the case of transformationally compound sentences like

(xiii)   The man who likes Mary is not the man who wrote
          the letter.

(xiv)   The woman who hates Peter is not the woman who got
          the letter.

though any native speaker of English would recognize (xiii) - (xiv) as being semantically related in some way.[12].

2.4 From a practical point of view, a variant of (iii) may be of more use. Namely, let us replace any category by a pair of numbers

(xv) $\pm (p,q)$,

where p stands for the level and q for the path, furthermore $1 \leq p \leq n$ and $1 \leq q \leq m$, where n stands for the number of levels and m for the number of paths.

This way (iii) is mapped into a finite subset of the infinite set of pairs of natural numbers. We obtain the following matrix:

$$(xvi) \quad \pm \begin{vmatrix} (1,1) & (1,2) & (1,3) & \dots & (1,q) \\ (2,1) & (2,2) & (2,3) & \dots & (2,q) \\ \dots \\ (p,1) & (p,2) & (p,3) & \dots & (p,q) \end{vmatrix}$$

Now the semantic characterization of a given word consists of a sequence of (xv) so that each number i for $1 \leq i \leq p$ occurs in the first place of (xv) at least once.

All definitions based on (iii) can be easily reformulated on the basis of (xvi). It is not a trivial consequence of this formulation that a similarity measure can be introduced which may be a useful tool in compiling thesauri or in language data processing. However, I shall not follow this line further at this place[13].

3. Contrast

3.1 We find the following definition of contrast in John
Lyons' recent book[14] (in a slightly revised form):
Two words, x and y, are said to be in contrast, if and only
if from x follows not-y and from y follows not-x but y need
not follow from not-x and x need not follow from not-y.
This definition is equivalent to the following one:
Two words, x and y, are said to be in contrast if and only
if both x and y belong to the same semantic class. By
semantic class we understand a set of words which may be
headed by a common word.

The antinomy relation is a special case of the contrast
relation. If from x follows not-y, further from y follows
not-x and vice versa, i.e. from not-x follows y and from not-y
x, then between x and y an antinomy relation holds.[15]

By way of illustration let us mention that "black" and
"white" would be in contrast because they may be headed by
the word "color", i.e. they belong to the same semantic class.
Or, alternatively, we could say that from "black" follows
"not-white" and from "white" follows "not-black" but not
vice versa, i.e. from "not-white" does not follow "black"
because it might be "red", "yellow" etc. and from "not-black"
does not follow "white" because it might be again "red",
"yellow" etc. On the other hand, however, if we take words
like "married" and "unmarried" or "sick" and "healthy" then,
obviously, an antinomy relation holds between them. The point
is that in the case of contrast the corresponding semantic
class contains more than two elements while in the case of
antinomy the semantic class contains just two elements.

No doubt, the only reasonable explanation for this
relation lies in the fact, that it has the same underlying
relation between categories. So "color" is a head word of
the words "white", "black", "yellow" etc. forming a semantic
class, because there is a subconfiguration of (iii)

$$c_k^i$$
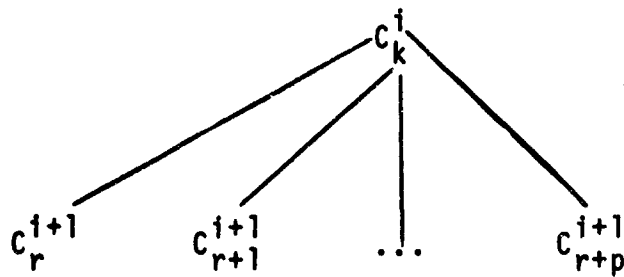
$$c_r^{i+1} \qquad c_{r+1}^{i+1} \qquad \ldots \qquad c_{r+p}^{i+1}$$

where the category c_k^i stands for "color" (maybe "color" as category) and c_r^{i+1} , c_{r+1}^{i+1} , ..., c_{r+p}^{i+1} for the corresponding categories for the different color names.

The categories c_{r+p}^{i+1} form a contrast set if $p > 1$ and an antinomy set if $p = 1$. It should be noted that it is by far not self-evident that the contrast set is finite. Let us consider, for instance, the following example:

(xvii)

shape

round oval ... triangular quadangular ...
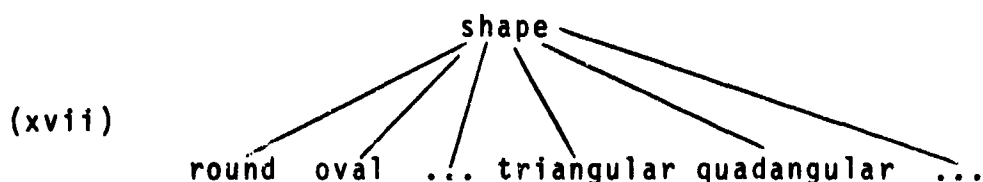
Notice, however, that "triangular", "quadangular", etc. are compound adjectives of a special kind, namely of type

n + angular

where n stands for any natural number. As a consequence we have to face here a syntactic problem and not a semantic one. We may, therefore, assume with good reasons that the contrast set of (xvii) contains "angular" instead of the infinite set of "triangular", "quadangular", etc. As a consequence we consider all contrast sets as being finite.

3.2 We may speak of various degrees of contrast as well. Let n_1 and n_2 be two levels of (iii). Further let us denote two different contrast sets by M_{c_1} and M_{c_2} belonging to the level n_1 and n_2, respectively. The contrast represented by M_{c_1} is greater than that represented by M_{c_2} if and only if

$n_1 > n_2$, it is of the same degree, if and only if $n_1=n_2$ and it is less, if and only if $n_1 < n_2$. So, for instance, the contrast is greater in the case of

$$\text{Being}$$

Human   Animal
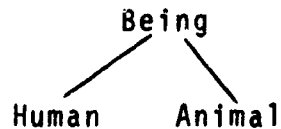
as in the case of (xvii).

3.3 There is a considerable difference between contrast and antinomy. This difference is brought to the fore by the effect of negation on both sets. For simplicity's sake let us denote the negation of x by $\bar{x}$. We have in the case of contrast the set

$$(x_1, x_2, \ldots, x_n).$$

If we say

Something is x_1

this means, it is $\bar{x}_2$, $\bar{x}_3$, ..., $\bar{x}_n$. On the other hand, if we say

"Something is $\bar{x}_1$."

then this means it may be either x_2, or x_3, or ... x_n.

In the case of antinomy, however, we have the set

$$(x_1, x_2)$$

and if we say

"Something is x_1."

that means it is $\bar{x}_2$ and

"Something is $\bar{x}_1$."

means it is x_2.

3.4 There is an apparent relationship between similarity and contrast/antinomy. It is clear that if x and y are two words in contrast/antinomy, then x and y are similar (because they share a head category) but not vice versa (because "man" and "boy", though similar, are not in contrast).

3.5 Contrast and antinomy may be useful by comparing sentences like

(xviii)   Ann is married.
            Ann is a spinster.

(xix)   It is a good book.
            It is an interesting book.

(xx)   The table is round.
            The table is rectangular.

I think it is necessary to differentiate between sentences like

            The green suit is black.
(xxi)   The long table is short.
            The old man is young.

and

            The man is a wife.
(xxii)   The bride is a groom.
            The winner is a loser.

Sentences like (xxi) and (xxii) are called contradictory sentences by Katz[16]. The contradiction in (xxi) however, is different from that in (xxii) and that is because the explanation for contradiction lies in the case of (xxi) in the fact that the "corresponding" words are in contrast while they are antinomous in (xxii). This gives sentences like (xxi) a different status from sentences like (xxii).

4. Inclusion

4.1 Let us take a sequence of words

(xxiii) $w_1, w_2, \ldots, w_n$

and the relation of (iv) between each (w_1, w_{i+1}) pair of
(xxiii). Let us further denote the set of meaningful sentences
by L. How it is true that the sequence (xxiii) may be
characterized by the following sentences:

$$w_1 \underline{is} \; w_2, w_3, \ldots, w_n \quad \in L$$

$$(xxiv) \quad w_2 \underline{is} \; w_3, w_4, \ldots, w_n \quad \in L$$

$$\ldots$$

$$w_{n-1} \underline{is} \; w_n \qquad\qquad \in L$$

and none of the sentences

$$w_i \underline{is} \; w_j$$

belongs to L where $i > j$.

As the relation (iv) has only been defined for categories
and not for words, we have to make an additional remark. If all
the words occuring as predicates in (xxiv), i.e. all w_i's except
for w_1, are categories and w_2 is a category of w_1 and finally
if the relation (iv) holds between pairs (w_i, w_{i+1}) for $1 < i \smile n$,
then (xxiv) is true and we may speak of an inclusion relation
between the words $w_1, w_2, \ldots, w_n$.

All that has been said is valid for any whole path of (iii)
(because each path will contain only categories between which
the relation (iv) holds).

4.2 By way of illustration let us take the following example:

        fox terrier, dog, mammal, animal

Then apparently,

        The fox terrier is a dog, a mammal, an animal.

        The dog is a mammal, an animal.

        The mammal is an animal.

all belong to L, but none of the following sentences belong
to L:

> The animal is a mammal.
> The mammal is a dog.
> The dog is a fox terrier.

These sentences are good examples of how relations between categories and relations between sentences interdepend.

4.3 It is again obvious that all words which have an underlying inclusion relation between the corresponding categories are at the same time similar as well (because they have at least one category in common, as, for instance, in the case of "animal" and "dog") but not vice versa (e.g. "man" and "woman").

5. In addition to the relation definded by (iv) and used in 4., which may also be refe.red to as <u>esse-relation</u>, there is anothe~ to some extent analogous semantic feature of natural language which may be called in contrast to the esse-relation <u>habere-relation</u>.

The problem will be clearer if we begin with the following sentences:

> The man has a head.
> The head has hairs.
> (xxv)　The head has ears.
> The ear has an earlobe.
> The head has a nose.
> The nose has a tip.
> The hair has a root.
> etc.

and on the other hand

> The man has a tip.
> (xxvi)　The man has a root.
> The man has a marrow.

While all sentences of (xxv) belong to L, none of (xxvi) will - at least under normal circumstances - belong to L.

This relation is one of the many relations which make it necessary to impose a more complicated structure on (iii)[17]. A hierarchy like (iii) would do only as a first approximation. However, it is not quite clear so far, what the structure of (iii) would be[18]. It seems as if we could account on the basis of (iii) only for <u>transitive relations</u>. Similarity, contrast, inclusion are apparently transitive relations, the habere-relation is, however, <u>intransitive</u>. We think that the latter are much more numerous in natural language. Let us point to the fact that, for instance, all verbs expressing a feeling toward another person represent an intransitive relation. Take, by way of illustration, the following example:

     (xxvii) Peter loves Mary.

             Mary loves John.

In fact, nobody would think that

             Peter loves John.

is a corollary of (xxvii).

    As I wish to tackle the question of intransitive semantic relations at length in a subsequent paper, I have to leave it with the above remarks.

6. To sum up, it seems clear enough to state that (iii) may be the basis for a definitional apparatus to be used in semantic analysis. Furthermore there are evident reasons in support of the claim that similarity as defined in 2. is a basic semantic relation and many others may be connected in one or another way with similarity. There are other language facts which suggest that neither the hierarchy (iii), nor the relations 2-4 are sufficient for the description of the semantic relations in natural language. On the other hand, it seems improbable that a system like (iii) can be set up for any natural language. What can be established is an <u>incomplete</u> system at best. As a consequence, the semantic characterization becomes incomplete as well. But even in the case of an

incomplete system (iii) the semantic relations as defined above may prove to be useful in semantic analysis. Further investigations are needed to decide this question[19].

Notes and References

1. Instead of "word" I would prefer the term "morpheme" - at least as far as agglutinative languages like Hungarian are concerned. Here, however, the term "word" refers simply to a lexicological unit.

2. It should be made clear that (ii) is not a formal requirement. It is, however, possible to define "relevantness" in a way that (ii) be - at least a semiformal requirement. Namely if we take any category of K, e.g. C_i, then C_i is linguistically relevant if there are at least two words, w_j and w_k, which are distinguishable just by the presence (or absence) of the category C_i. To render (ii) totally formal we would have to define the semantic characterization of words in a non-trivial way.

3. A more detailed discussion of these questions is to be found in Kiefer-Abraham.

4. As far as I know a system like (iii) has first been proposed by Chomsky. Cf. Chomsky 1961.

5. So, for instance, the category "Abstract" would occur both in the characterization of "love" and "to love".

6. It would be impossible to compare words like "cash" and "to cash" because - as it follows from the nature of any hierarchy - the comparison should begin "on the top", i.e. comparing categories belonging to the highest level and then proceeding downwards. Words like "cash" and "to cash" would apparently differ already on a very high level (probably already on the second level) and as a consequence the comparison procedure would be blocked.

7. Inclusion as a general property of language has been described - from various points of view - by Chomsky (syntax), Katz and Postal (semantics) and Bierwisch (lexicology. Cf. Chomsky 1965, Katz-Postal and Bierwisch.

8. The relation (iv) implies that no category may occur twice.

9. The labelled tree form for the semantic characterization
   of words has been proposed by <u>Katz-Fodor</u> <u>1963</u>. However,
   the "normal form" of Katz-Fodor has turned out to be
   unsatisfactory for many reasons as has recently been
   pointed out by <u>Weinreich</u>. Almost the same could be said
   against my proposal but I c not consider it more than a
   starting point. It is e .. possible that the semantic
   characterization will be so complicated that tree struc-
   ture will no longer be able to visualize it.

10. Similarity as a basic semantic relation but defined in
    a different way has been treated at length by <u>Spark</u> <u>Jones</u>.

11. This equivalence relation leads to a partition of the
    vocabulary. Each class will contain a stock of "similar"
    words.

12. Transformational grammar could help in this respect but
    so far we do not know very much about transformation either.

13. Some practical work has already been done in this direction.
    (See the forthcoming issues of <u>Computational</u> <u>Linguistics</u>.)
    From a theoretical point of view <u>Brodda's</u> paper is worth
    mentioning.

14. Cf. <u>Lyons</u>.

15. These definitions should refer to one meaning (i.e. to one
    path on the tree diagram) of the words x and y. Since we
    do not make use of this restriction here we leave it out
    of consideration.

16. Cf. <u>Katz</u>.

17. For a more detailed treatment of this topic see <u>Bierwisch</u>.
    The failure to explain the intransitivity of the hab e-
    relation lies in the fact that no hierarchical system of
    type (iii) can account for such relational terms as "tip",
    "root", etc.

18. (xvi) could be imagined as consisting of n-tuples instead
    of pairs of numbers. That means that the corresponding
    (or underlying) tree representation would be n-dimensional.

19. The best proof that such a system may be useful is John
    Lyons' book (see <u>Lyons</u>). Some additional questions of
    a formal semantic theory are tackled in Kiefer and
    Abraham-Kiefer, although both of these papers cannot be
    considered more than a very tentative approach to the
    semantics of natural language.

BIBLIOGRAPHY

Abraham, S., F. Kiefer. A Theory of Structural Semantics,
        Mouton & Co., forthcoming.

Bierwisch, M. "Eine Hierarchie syntaktisch-semantischer Merkmale,"
        Studia Grammatica V., Berlin, 1965.

Brodda, B. A Measure for Similarity, KVAL Papers 1965,
        Stockholm (mimeographed).

Chomsky, N. "Some Methodological Remarks on Generative Grammar,"
        Word 17, 1961, pp. 219-239.

_____. "Categories and Relations in Syntactic Theory,"
        M.I.T. Cambridge, Massachusetts, 1964 (mimeographed).

Katz, J.J. "Analyticity and Contradiction in Natural Language,"
        The Structure of Language (eds. J.A. Fodor & J.J. Katz),
        Prentice-Hall Inc., Englewood Cliffs, N.J., 1964,
        pp. 519-544.

Katz, J.J. J.A. Fodor. "The Structure of Semantic Theory,"
        Language 39, 1963, pp. 170-210.

Katz, J.J. P.M. Postal. An Integrated Theory of Linguistic
        Descriptions, M.I.T. Press, Cambridge, Massachusetts,
        1964.

Kiefer, F. Questions of a Formal Semantic Theory (in Hungarian),
        forthcoming.

Kiefer, F. S. Abraham, "Some Questions of Formalization in
        Linguistics," Linguistics 17, pp. 11-20, 1965.

Lyons, J. Structural Semantics, Philological Society, Oxford,
        1964.

Sparck Jones, K. "Synonymy and Semantic Classification," Cambridge
        Language Research Unit, M.L. 170, 1964.

Weinreich, U. "Explorations in Semantic Theory," forthcoming in
        Current Trends in Linguistics, vol. III (Th. A. Sebeok,
        editor).

DISCUSSION

ULLMANN: These categories: Did you say that you are skeptical about applying them to the more complex cases? I think they can be. Lyons, in his Structural Semantics, has a very similar set. There are one or two which you didn't mention but which are really derivatives, the complementary ones, like "buy" and "sell."

He seems to have handled the whole corpus of this rather complex semantic field quite successfully in terms of these half-dozen features.

KIEFER: I think so too, but what I don't know so far is the problem of idioms and stylistic problems and so on. But I think it is not proper to exclude it from semantics. It may be considered as a tool for describing semantics.

BAR-HILLEL: How are you going to handle, with the help of any semantic categories, things such as "A is a point between B and C, and if A is between B and C, then A is between C and B?" Take this example of a meaning rule, that "If A is between B and C, A is between C and B." Do you really envisage that you are going to handle this in some kind of category?

KIEFER: No. What I think is, it may be handled in terms of categories. I mean, if you think in terms of the Katz-Fodor theory, this is the lexicon, and between the projection rules you have to introduce a definition apparatus, something like a similarity or contrast, and a lot of other relations, and probably include not only categories like commonplace categories, as "human being," but you can even give categories which give direction, or something like that, and apply a different kind of handling.

ULLMANN: Katz himself is very interested in field properties, which I know from personal conversation.

MASTERMAN: What is the difference between a field category and a property?

ULLMANN: Field properties are these organized lexical sectors, like the aforementioned "color," or Lyons' intellectual fields. Chomsky's point in aspect is -- but he makes his very briefly -- that the Katz-Fodor semantic markers don't exhaust all there is to be said on the meaning of these words in the dictionary part of the semantic part of the generative grammar, but how the field properties can be assimilated into the scheme or added onto it he doesn't say, and we don't know.

MASTERMAN: What does "field" mean?

BAR-HILLEL: Some schools call it "lexical field."

ULLMANN: An organized sector of the vocabulary --

BAR-HILLEL: For which thesaurus is a close approximation.

SPARCK JONES: I made my comment in saying it, that these semantic fields, if they are anything like thesaura, they are defining categories. You can't say they are quite different.

ULLMANN: It is not a question of "green" belonging to the category of "color." It is placed in that category, in that particular category. It is not specifically a category exclusion rule.

BAR-HILLEL: For instance, "Orange is between yellow and red."
This can not be handled under "color." Orange in ⌐ very
important sense is between yellow and red.

ULLMANN: Take for example "father." It is not enough to say
it has a certain point in the hierarchy.

VON GLASERSFELD: I think the tone and the way the explosion
"first approximation" came out a moment ago is indicative of
something that has been boiling under the surface of this
meeting all along. There are two kinds of people here; the
ones who would like a theory of semantics that embraces abso-
lutely everything that can be done with language, and there is
another kind here who will be very happy to have any kind of
first approximation that works in any little field of semantics.

I think this is a distirction that is traditional and,
as I said to someone before, it reminds me of doing chemistry
before Mendeleev and his periodic system. There is no question
that chemistry was better afterward, but some of the chemistry
done before was pretty good.

VIII.

UNDERLYING STRUCTURES IN DISCOURSE

by

Thomas G. Bever and John Robert Ross
MIT and Harvard University

In this paper, we will address ourselves to several
semantic problems which arise in the attempts to give a
precise characterization of the properties that make a
sequence of sentence into a coherent text. But these
problems are of such complexity and depth that we will not
be able to present a solution to any of them. It is,
however, our belief that they have a crucial bearing on
various aspects of semantic theory, and we hope that
presenting them here will serve to redirect attention to
areas of investigation which have been neglected of late.

Consider first the problem of the semantic interpre-
tation of discourses. Clearly, any adequate theory of
semantics must somehow express the synonymy of (1) and (2):

(1) A bullet will kill a pullet.

(2) a. Something will happen to a chicken.

    b. The chicken is young.

    c. Something will cause the chicken to enter a state.

    d. The state is death.

    e. The instrument of the change of state will be a
       bullet.

Katz and Fodor[1] propose to interpret a discourse by
conjoining all its sentences and applying the semantic rules
to the result. They say (p. 491)

> "Hence, for every discourse, there is a single
> sentence which consists of the sequence of n
> sentences that comprises the discourse connected
> by the appropriate sentential connectives and
> which exhibits the same semantic relations exhi-
> bited in the discourse."/emphasis ours -T.G.B.
> and J.R.R./

If Katz and Fodor mean the resulting conjoined sentence
to have a coordinate structure, with all the original sentences
of the discourse dominated immediately by the same node S,
then their proposal would seem to be clearly wrong, for it is
a commonplace that some sentences in a discourse are more
closely semantically related than others. For instance, (2a)
and (2b) above are more closely related than either is to
(2d). A coordinate conjunction of the five sentences (2a) -
(2e) would obscure this fact. But, if Katz and Fodor are
taken to be asserting that it is possible to preserve the
semantic relations among the sentences (2a) - (2e) by forming
some kind of non-coordinately conjoined sentence, then they
are simply begging the question.

A proposal which seems at first to be more promising is
the following: in interpreting a discourse, we will replace
each anaphoric expression (i.e., pronouns; determiners like
the, other, this, such, etc.) by its full antecedent and
and then use the resulting sentences as the input to the
semantic rules. Thus (2b), (2c), and (2d) would be replaced
by (3b), (3c), and (3d):

   (3) b. The chicken to which something will happen is young.

       c. Something will cause the young chicken to which
          something will happen to enter a state.

       d. The state which something will cause the young
          chicken to which something will happen to enter
          is death.

So far so good. But notice that there is no simple way
of finding the full antecedents of the instrument and the
change of state in (2e). Even if some fairly reasonable
solution can be worked out for this case, we believe that
the general problem has no easy solution. Notice also that
this method misses an important semantic relationship between (2a)
and (2c): the fact that the verb cause to enter a state
is a "happening" verb. The sentences (2a) - (2c) would not
form a discourse if (2c) were replaced by (2c'): (2c') the

chicken will appeal to you. The reason for this is, of course, that the verb appeal to is not a "happening" verb.

It seems, thus, that this second proposal will not work either. What seems to be necessary for us to be able to mark (1) and (2) as synonymous is some more abstract structure which would underly both. We will speculate briefly on the nature of such a structure below, after we have discussed the two main properties of discourse.

Following Lakoff[2], we will say that to comprise a discourse, a sequence of sentences must be connected and structured. A set of sentences is connected if all share a sufficient amount of semantic material. Just what constitutes "a sufficient amount" is a difficult question, to which we will return below. A sequence can only be structured if it is connected, but the converse is not true. An example of a connected but unstructured text may bring out the differences between connectedness and structure:

(4) a. It takes a month's wages to buy a pair of shoes in Russia.

   b. Russia was once ruled by tyrannical czars.

   c. Tyranny is almost always overthrown by a revolution.

   d. The American Revolution started when a minuteman fired on a redcoat.

Although the sentences in (4) are pairwise connected by the concepts Russia, tyranny, and revolution, no discourse results, because the topic changes from sentence to sentence. However, it is interesting that the sentences in (4) can begin a discourse, if we add such sentences as (5) to them.

(5) a. This shot touched off a bitter conflict which was largely caused by the oppressive laws imposed on British colonies by King George III.

   b. The American victory can be attributed to the wide popular support which the leaders of the rebellion had.

   c. In a bloody insurrection in 1917, Russia's nobles were either murdered or forced to flee the country by a huge peasant population suddenly gone amok.

d. But it is one thing to rid a country of an oppressive
system, and another to provide it with a strong
economy. Since 1917, Russia has been engaged in a
grim struggle for economic survival, but today's
living standard is only slightly better than that of
1917.

We claim that, while (4) - (5) is not felicitous, it is
still a discourse. The impression one gets when reading it
is that one is reading a complicated formula, which starts
out with a lot of left parentheses, or that one is hearing
a self-embedding sentence like (6) or (7).

(6)   That that that that he came surprised me was amusing
to her was obvious is possible.

(7)   A boy who a man who a book which I read fell on was
cursing at ran away and hid.

Notice that the order of the sentences in (4) - (5) is
strictly fixed: if (5c) followed (5d), the sentences would
no longer form a discourse. The sentences in (5) are linked
to those in (4) in reverse order: (5a) is linked to (4d) by
the phrase this shot, and to the word tyranny in (4c) by the
phrase oppressive laws. (5b) is linked to (4c) by the word
pairs victory - overthrown, revolution - rebellion. (5c) is
linked to (4b) because both are about Russia, and (5d) is
linked most strongly to (4a).

This example suggests that discourse "structure" may,
at least in some cases, be attributed to the operation of a
recursive discourse formation rule. Here, one such rule
extends a well-formed discourse by inserting a well-formed
sub-discourse into it at some point, subject to restrictions
on connectedness. For instance, the sentences (4a) - (4c)
and (5c) - (5d) constitute a discourse in themselves. The
sub-discourse (4d) - (5a) - (5b), which is about the American
Revolution, can be inserted after sentence (4c), because it
is connected to it by the word revolution. At present, we
do not know to what extent intuitively felt "structure" in
other kinds of "connected" sentence sequences will be able

to be accounted for on the basis of discourse formation
rules like the one sketched above.

Let us return now to the notion of "connected" sentences.
Above we asserted that sentences are only connected if they
share "a sufficient amount" of semantic material. This
proviso is necessary, for surely we would not wish to assert
that sentences (8) and (9), which share only the marker
(Physical Object), are connected:

(8) This car sure runs well.

(9) Tom ate a snake.

This example indicates that there is some lower bound on
connectedness, although we have no idea at present of how
to characterize it. But it will almost certainly depend
not on the number of shared semantic properties, but on
their kind. It may be that features like (Physical Object),
which never contribute to connectedness, can be formally
distinguished on independent grounds from those features
which do contribute to connectedness.

Notice that it is not the case that sentences are only
connected to their neighbors. In fact, if the sentences in
(5) above were only connected in this way, (4) - (5) would
not be judged to be a discourse. For instance, the inserted
sub-discourse (4d) - (5a) - (5b) is linked by pairs oppresive -
tyranny, victory - overthrown, revolution - conflict, wide
popular support - huge peasant population, conflict - struggle,
rebellion - insurrection, victory - survival, etc. to every
 ther sentence in the text except (4a). This means that the
discourse formation rule discussed above must be restricted
in some complicated and non-obvious way so that an embedded
discourse will be required to tie in with the whole surrounding
text, not just its nearest neighbors. For otherwise, the
insertion of a sub-discourse which was not "multiply connected"
would destroy the coherence of the whole discourse.

In conclusion, we would like to raise the question of whether it is likely that underlying structures for discourse, whatever they turn out to look like, can be generated by a device that has no access to extralinguistic material. In this light, consider the discourse (10) - (11):

(10) I had an accident in my car yesterday.

(11) The right front fender is totally ruined.

Conceivably, one might argue that this discourse is elliptical, and that the interpretation should not take (10) - (11) as input, but rather (10) - (10a) - (11), where (10a) is

(10a) Cars have fenders.

It can be shown that such have a sentences as (10a), which express an inalienable-part hierarchy, are necessary to derive only grammatical English sentences, so one might argue that such sentences are available in forming discourses by elision. But how could we construct a similar argument for a discourse where (10') replaces (10)?

(10') I had an accident while driving yesterday.

To take a more extreme example, consider the discourse (12) - (13).

(12) I think you should take a look at the Bible.

(13) The Ten Commandments have been an inspiration to young and old readers for centuries.

If the phrase the Ten Commandments in (13) is replaced by Gödel's Incompleteness Theorems, the sequence of sentences ceases to be a discourse. And clearly the fact that the Ten Commandments are in the Bible, while Gödel's theoremes are not, is not a linguistic fact. Similar examples are not difficult to construct.

To us, these facts seem to indicate that the search for underlying discourse structures within the bounds of linguistics is futile. Rather, what seems to be necessary is some kind of concept generator which, having access to our entire belief and concept networks, produces some kind of abstract object which represents the maximal content of

a whole set of discourses which derive from this concept. Then
some kind of mechanism must select certain aspects of this
abstract object which are to be communicated and somehow select
lexical material to accomplish these ends. In the process of
selection. a speaker clearly estimates the previous knowledge,
beliefs, and reasoning power of his audience, and leaves parts
of the concept unexpressed, on the assumption that the audience
will be able to fill them in. In other words, we would say
that the discourses uttered in response to the question, "what's
a carburetor?", whether in answer to a question asked by a
five-year-old boy or by a twenty-year-old man, have the same
underlying structure, despite the fact that these discourses
will differ in radical ways. The linguistic meaning of each
of these discourses is now only a part of the entire concept -
the part that in each case has been put into words.

It should not need to be emphasized that the above
proposals are highly speculative, and that we have no idea
about how to go about implementing them. Nevertheless, we
feel that only a device with access to extralinguistic material
can explain the notion of connectedness in discourse.

In summation, we have suggested that while it may be
possible to state discourse formation rules which provide an
account of structure in discourse, the problem of connectedness
in discourse cannot be solved within the confines of linguistics.
Since the problem of interpreting discourses by semantic rules
clearly presupposes the establishing of the correct connections
between parts of discourses, the problem of semantic inter-
pretation of discourses is also unsolvable within linguistics
proper.

Acknowledgement

FOOTNOTES

1. Cf. p. 490-491 in Jerrold J. Katz and Jerry A. Fodor "The Structure of a Semantic Theory." _The Structure of Language: Readings in the Philosophy of Language_, Katz and Fodor (ed.), Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1964.

2. Cf. George P. Lakoff, "Structural Complexity in Fairy Tales," unpublished mimeograph, Indiana University, January 1964.

DISCUSSION

HAYS: This brings out the very interesting fact that whereas parsing is a natural part of a performance model, generation production is not. That is, parsing is part of a recognition procedure, a part that would come naturally before concentratic of cognitive networks and all that you know, whereas the generation of a grammatical structure is not a natural part of the production of a sentence when you have begun with some structure of cognitives of some kind. What you need in that part of the performance system is a lot of transduction. This is pretty close to the Leroy model -- André Leroy.

ROSS: What was the example?

HAYS: What he proposed was the storage of a great network of factual knowledge which would be developed from the analysis of documents by an automatic procedure that would be grammatical and semantic and rely on as much of that network of factual knowledge as can be developed today. That is, the linguistic system would have at one end parsers and sentence manipulators, and on the other side a kind of cognitive network. This, it seems to me, is a substantially different proposal for abstract semantical systems than the one of attributing properties to things, since in the underlying network there can be two-place predicates and three-place predicates and as much complexity of that sort as is required.

ROSS: I might add that it is quite possible that we will still be able to do semantic analysis of sentences within the bounds of linguistics, because really the property of connectedness is orthogonal both to grammaticality and to semantic well-formedness. The sentence "I saw a whale yesterday and 2+2=4" is grammatically well formed and presumably on some level also

semantically well formed. However, it is not connected.

You see, the problem of connectedness does not really raise its ugly head with full force until you actually try to separate sequences of sentences which aren't discourses from those which are. Then you must have connectedness; otherwise you have nothing.

MASTERMAN: I don't know; I think I really do disagree with your extreme gloom. I find it a little difficult to see why. Will you let me take your example about the Bible?

My underlying feeling is we have two quite different notions of "meaning rule". I can't find yet in what they are different, but if you take this example about the Bible, suppose I didn't know the Ten Commandments were in the Bible? I would nonetheless infer, simply from the concatenation, that they were.

Supposing we do put in Gödel's Incompleteness Theorems and suppose I don't know, really, what the Bible is, but it's clear by what comes after "the Bible", simply by the position in the discourse, is going to be connected to it, and perhaps I don't even know the Bible is a book. It is still the case that I think I could make a machine infer that Gödel's Incompleteness Theorems were in the Bible, and this would be a wrong fact that it recorded. Nevertheless, it would be a wrong fact that it recorded, and normal discourse that gets understood doesn't, as you have said, record these wrong facts. I mean, we do get to know things from discourse that we didn't know before, and the kind of rule, meaning rule, that gets you to know something because it's said to you, because you know so much about the positioning of the important words in what is said to you in very much the way you gave, is a different conception of meaning rule from the kind of meaning rule that says, "I would have thought a rule of physics" --

BAR-HILLEL: Don't call it meaningful. It is perfectly all
right. It is the tendency of human beings to impose discourse
structure even to something which at first sight doesn't have
anything. This is perfectly all right, because you know what
other people are saying.

MASTERMAN: I way trying to illuminate the notion of connect-
edness. I was trying to elucidate your notion of connectedness,
which I am sure is cardinal, by giving rules of connectedness,
if you like, that a machine will pick up. We listen in order
to learn things. How do we learn them? Because we are listen-
ing for something.

CHARNEY: I was sort of on David Hays' side. Why do we have to
generate connectedness? After all, isn't this the kind of
thing that the human being uses the language for? He uses the
language; he knows the ordinary rules, he knows what comes
close. There are certain rules of juxtaposition, certain rules
of reference that go out beyond. Nevertheless, words like
"nevertheless," "however," "anyway," and so on, go far beyond
the confinement to a single question; in a connected discourse
there are no bounds. You can refer back over a thousand years.
So why and how would it even be possible to say that we can
not solve this problem of connected discourse in linguistics
simply because it is impossible to generate connected discourse?
No mechanism decides what is relevant. You have essentially
a discourse form which is very, very abstract.

ROSS: I think that we are in complete agreement. However,
what seems to you to be obvious apparently has not seemed to
people like Harris to be so obvious, because Harris has tried
to construct a set of rules for establishing discourse
connectedness, essentially, which are not even semantic. I

take it that Harris would like to make the assertion that
by and large the connections in discourse are not even semantic.
You don't even need semantic knowledge to connect discourses,
and you can do pretty well just with a grammatical equivalent.
If it is an open and shut issue to you that discourse is not
semantic and not linguistic, then fine. Then I have said
nothing new.

CHARNEY: I didn't say it wasn't semantic. Of course it is
semantic. But the thing is, every time I say something I get
new information, and I have a purpose behind my connecting
things that have not been connected before, so I can't put
restrictions on what possibly can be generated. I am not a
mechanism generating one sentence after another. I am a
thinking human being using very well-known rules of language
that all of you know and all of you understand, and in this
way you understand the import of the total effect of every-
thing that I am saying.

ROSS: Well, I guess we're in disagreement, then, on one
point; i.e., I would disagree that it is a semantic fact
about English, or that "The White House has a Blue Room"
that "My office is in Building 20 in MIT"; that "The Bible
contains the Ten Commandments." I would say any of these
facts can be used to connect the discourse. They are not
semantic facts; they are facts about the real world.

CHARNEY: That is why language is used as a communication
about the real world.

IX.

MULTIDIMENSIONAL SCALING AND SEMANTIC DOMAIN

A. Kimball Romney
Harvard University

This paper represents some preliminary results of continuing research, the major goal of which is to explore some ways in which semantic domains vary in internal structure.

For present purposes, a semantic domain may be defined as an organized set of words (or unitary lexemes), all on the same level of contrast, that refer to a single conceptual sphere. The words in a semantic domain derive their meanings, in part, from their position in a mutually interdependent system reflecting the way in which a given language classifies the relevant conceptual sphere. This definition corresponds to what Conklin calls "the basic level of contrast" (1964, p. 39) and to the notion of "lexical field" as used by Ohman (Word 1953).

In a recent article, Berlin and Romney (1964) gave the following example:

> An example of a semantic domain in English is "shape." Thus, words such as "round," "square," "rectangular," etc., may each be thought of as sharing the feature of "saying" something about shape. They signal the hearer that the aspect being talked about is shape. In addition, each word in the domain "says" something different, e.g., round is different than square. "Shape" is the gloss for a semantic domain or category. "Round," "square," etc., are members of the category.

Other examples of semantic domains include color terms, names of the months, kinship terms, names of the letters of the alphabet, disease names, plant names, pronouns, etc.

In this paper the primary interest is in making inferences about structure from the "distance" among items in the semantic domain. The methods that we use are not typically employed by linguists and are thought of as complementing more traditional linguistic methods. The methods are essentially those of scaling and involve attempts to measure distance among the words in a semantic domain by the method of judged similarity. Inferences concerning the structure are then made from the estimates of distance.

This method of arriving at the internal structure of a semantic domain provides an independent measure from that reached by the linguistic method. Thus, a structure arrived at on the basis of purely linguistic criteria may be compared to the structure arrived at on the basis of scaling methods.

So far in our research, we have isolated four major types of structure exhibited by various semantic domains. These are:

    I.   Scales
        A. Unidimensional
        B. Multidimensional
            1. Closed (circumplex)
            2. Open
   II. Taxonomies
  III. Paradigms
   IV. List structures
        A. Closed
        B. Open

Let us discuss briefly the characteristics of each of these major types.

Scales. For the sake of convenience, scales have been subdivided into unidimensional and multidimensional types. The sociological and psychological literature contains many examples and discussions of the unidimensional scales. They generally measure some characteristic or quality in a single dimension. We will not discuss them further here, although it should be

pointed out that a great number of simple semantic domains
may take the form of being ordered in the form of a unidimen-
sional scale. For practical purposes, our discussion of multi-
dimensional scales will be limited to two dimensions. We shall
see later that multidimensional scaling techniques in more than
two dimensions may take the form of paradigms or taxonomies
(paradigms and taxonomies, of course, may occur in two dimen-
sions). One common form of a two-dimensional scale is what
Guttman has called "a circumplex structure" (1954). We have
labeled these "closed" structures. His notion is that quali-
tatively different traits in a given domain can have an order
among themselves without beginning or end.

In order to illustra   a domain that exhibits this struc-
ture and to explicate our methods, let us consider for a moment
the domain of color in English. Utilizing a multidimensional
scaling technique described by Torgerson (1958, chapter 11),
we collected data on six common English color terms fr sixty
college students. The technique is called the triad method.
The six color names are arranged in all possible triads and
presented to the subject who is instructed to circle the
name that is most different of the three.

> This technique produces a distance model consisting
> of a set of absolute distances (of undetermined
> units) between all pairs of stimuli in the universe
> treated. These distances give the relative location
> of the stimuli in an n-dimensional space -- where
> n is the minimal number of dimensions needed to
> define uniquely the geometrical model. It does not
> yield a spatial model, e.g., it does not give the
> absolute projections of each point on axes referred
> to a known origin. The distance model is sufficient
> for our purposes, however, since we need only know
> the distances between points, and not their absolute
> locations in the n-dimensional space (Romney and
> D'Andrade, 1964).

Table 1 presents the innerpoint distances among the six
colors. The best geometrical representation is presented
in Figure 1. Note that the smallest distances are between
adjacent colors in the figure and that the greatest distances
are between colors on opposite sides of the figure with inter-
mediate distances among colors separated by one other color.

Figure 1.   Best Geometrical Representation of
             Interpoint Distances for Color Terms.
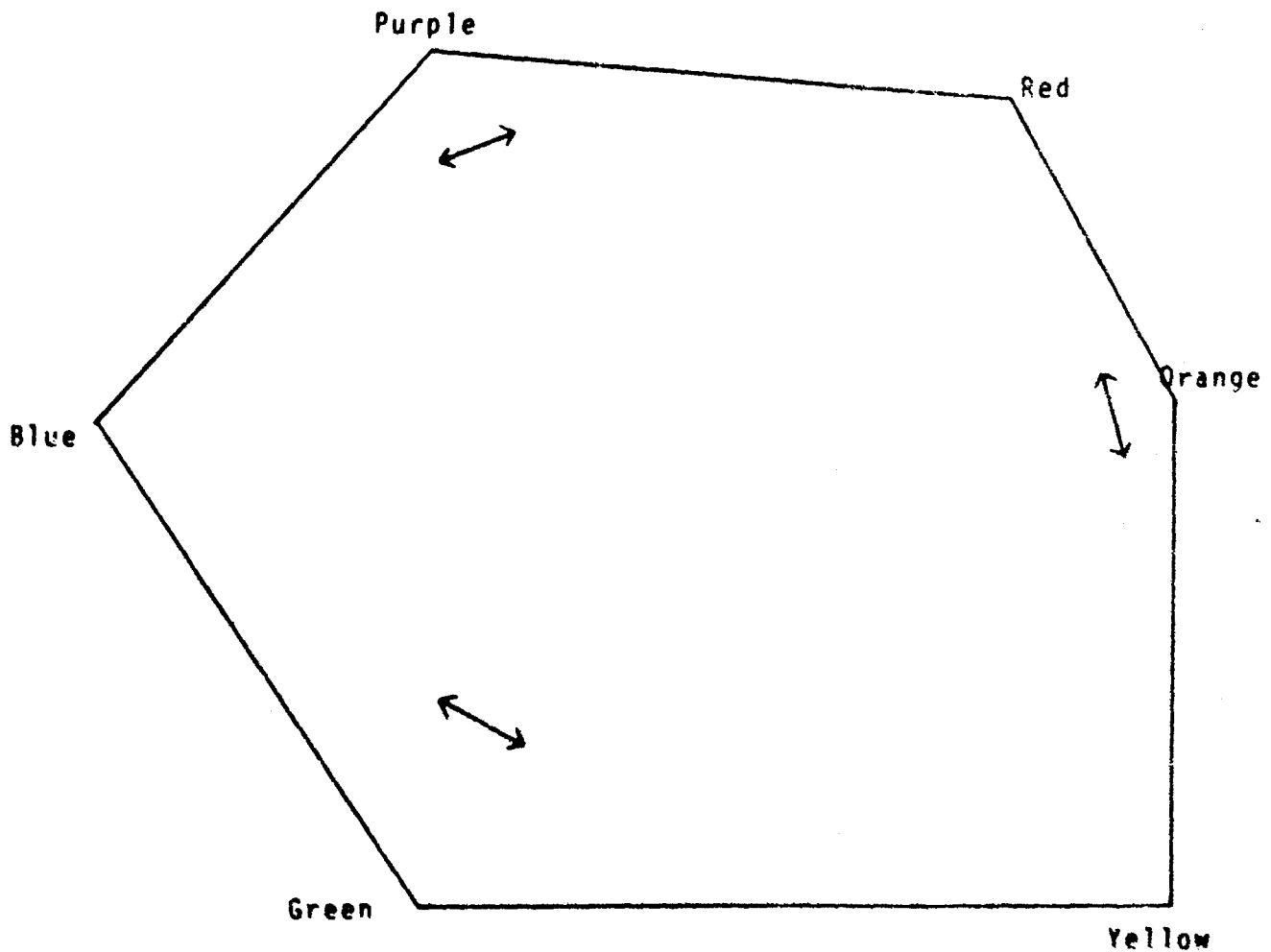
Table 1. Interpoint Distances among Six Color Terms for
60 Subjects.

	Red	Orange	Yellow	Green	Blue	Purple
Red	x	1.9	3.5	5.6	3.6	3.2
Orange		x	2.6	4.9	6.1	4.2
Yellow			x	4.3	5.3	6.4
Green				x	3.2	5.4
Blue					x	2.6
Purple						x

Color perception, of course, has been well studied by the psychologist, and it is no surprise that a circumplex structure should emerge utilizing a simple scaling technique on the color names. In considering closed multidimensional scales, the critical criterion is the closed sequence of variables. Absolute circular form is not necessary.

The second form of multidimensional scaling, the "open," lies somewhat between unidimensional scale and a circumplex structure. An example of such a structure from our own work has to do with the semantic domain of personality trait names, such as rude, bold, etc. Table 2 and Figure 2 present analyzed triad data on a sample of eight such names collected from sixty-three college students. Note that the geometrical representation does not "close" as for the color terms. The scale has a clear cut beginning and end point that requires at least two dimensions for its representation. It is generally crescent shaped, although the family of scales may take a variety of forms.

Figure 2. Best Geometrical Representation for Interpoint Distances for Personality Trait Names.
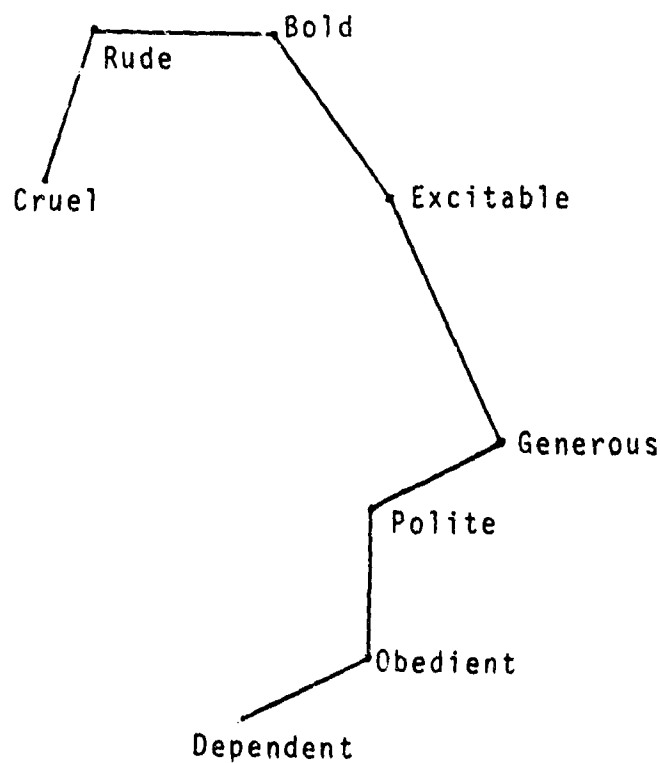
Table 2. Interpoint Distances among Eight Personality Trait Names for 63 Subjects.

	cruel	rude	bold	excitable	generous	polite	obedient	dependent
cruel	x	.8	1.4	1.9	2.9	2.9	3.1	3.0
rude		x	1.0	1.7	2.8	2.2	2.9	3.1
bold			x	1.2	1.5	1.9	2.3	2.8
excitable				x	1.6	2.3	2.6	2.8
generous					x	.8	1.4	2.0
polite						x	.8	1.3
obedient							x	.8
dependent								x

Taxonomies and paradigms. A taxonomy is generally
thought of as a tree structure in which the distinguishing
features in the various branches are different, one from
another. In distinguishing a taxonomy from a paradigm, we
follow the definition of Lounsbury:

> In the perfect paradigm, the features of any
> dimension combine with all of those of any other
> dimension. In the perfect taxonomy, on the other
> hand, they never do; they combine with only one
> feature from any other dimension. In the perfect
> paradigm there is not hierarchical ordering of
> dimensions that is not arbitrary; all orders are
> possible. In the perfect taxonomy there is but one
> possible hierarchy. To illustrate the difference
> we may consider a set of eight elements consti-
> tuting a field F. If these represent a paradigm,
> it takes but three dimensions of dichotomous
> opposition to fully characterize them (Figure 1).
> If they represent a taxonomy, it takes seven
> (Figure 2).

When utilizing distance data only, how does one distinguish
between a taxonomy and a paradigm? The answer to this question
is very clear cut.

In Figure 4a a simple paradigm is illustrated. In such a
structure, A is closer in distance to B and C than to D. In
the taxonomy of Figure 4b, A is closest to B and equidistant
to C and D. It is therefore possible to make inferences about
whether objects such as A, B, C, and D form a taxonomy or
a paradigm by an examination of the interpoint distances
among the objects or words.

As Lounsbury says,

> Kinship terminologies usually represent something
> intermediate between these, the imperfect or asym-
> metrical paradigm, which combines principles of both
> kinds. In the analysis of content fields other than
> kinship, one must be prepared to find both kinds of
> structures. Anthropological work on folk taxonomies
> reckons with both.

Figure 3. (Lounsbury's figures 1 and 2)

fig. 1   fig. 2

Figure 4. Paradigm and Taxonomy.
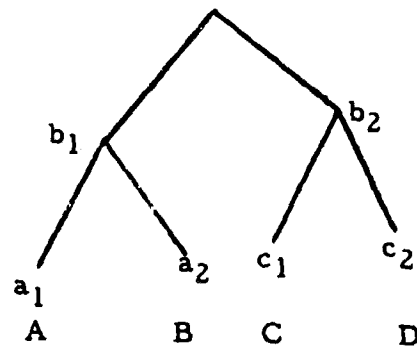
4a.   4b.

In our own work, we have not isolated any structures
that approach an ideal taxonomy. English kinship terminology
approaches a paradigmatic structure. Table 3 and Figure 5
present the data on the triads test for male English kin terms.

Figure 5. Best Geometrical Representation of Interpoint
Distances for Male Kin Terms.

Table 3. Interpoint Distances among Male Kin Terms for
          64 Subjects.

	GrFa	GrSo	Fa	So	Br	Un	Ne	Co
GrFa		2.44	2.20	3.35	4.12	3.55	4.60	4.60
GrSo			3.25	2.12	3.92	4.66	3.44	4.41
Fa				2.25	3.24	3.03	4.56	4.84
So					3.13	4.79	3.26	4.15
Br						3.28	3.55	3.26
Un							2.63	2.70
Ne								2.68
Co								
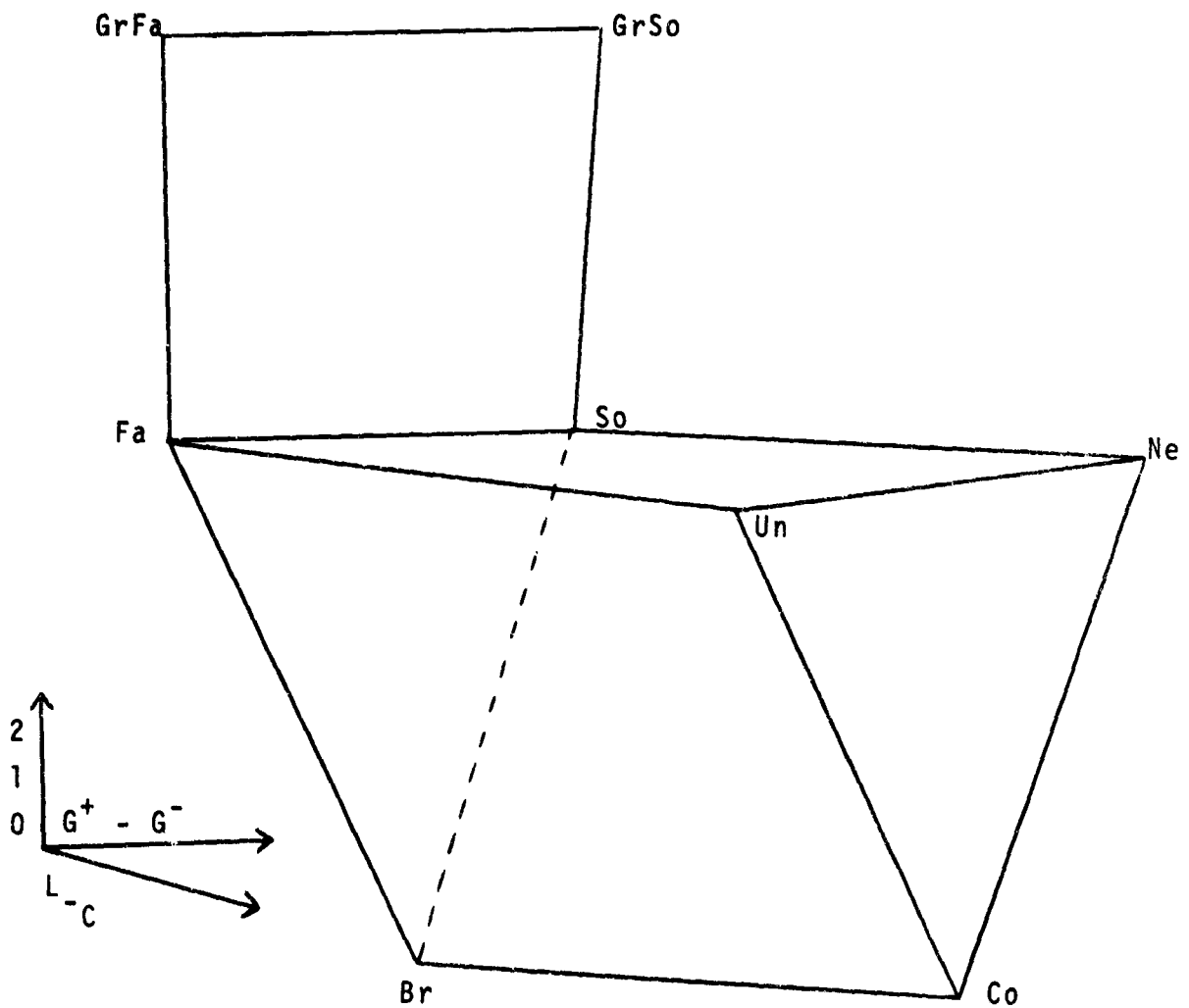
List structures. List structures may be thought of as highly internalized and ordered names of objects within a semantic domain. In a certain sense, they are "weak" scales. The days of the week, for example, or the names of the months seem to be closed list structures. The letters of the alphabet would seem to be an open list structure with a definite beginning and end.

Conclusion and discussion. In conclusion, I would like to make explicit some of the implications of the above discussion. First, we feel that different semantic domains may exhibit quite different structures. We feel that it would be a mistake to attempt to force all domains into taxonomies or paradigms.

Second, we feel that measures of similarity as represented in the triad text add information that is complementary to information arrived at by more strictly linguistic methods.

Third, though we have not mentioned it explicitly above, it is quite clear that each method imposes restrictions upon the types of results possible. The triad test is only an example, and we should seek other methods for studying the structuring of semantic domains.

Fourth, in our own work, we have found a fair amount of variability from individual to individual in the structuring of semantic domains. The most variability has occurred with a more complex structure, such as a paradigm.

I would like to expand on two of these points. The first is that various semantic domains exhibit different structures. A typology of some common structures is suggested. Second is that various methods of analysis should be applied to the same semantic domain. Each method adds to the total amount of information concerning a given conceptual area. Several methods taken together will also frequently determine which of various alternative formal analyses are most productive.

We can illustrate the results of various techniques and how they reinforce one another with the semantic domain of English kin terms. For this illustration, we deal only with the lineal terms. One way of partitioning a subset is on the basis of their occurrence with various modifying terms. Table 4 presents the results of the co-occurrence of kin terms together with the more common modifiers. Figure 6 shows the partitioning on the basis of similar patterns of occurrence.

Table 4. Percentage of Subjects Modifying Lineal Kin Terms with Common Modifiers (frequency below 10 excluded) N = 105

	step	in-law	great	half
Father	55	54	---	--
Mother	55	57	--	--
Son	20	28	--	--
Daughter	20	30	--	--
Brother	55	73	--	28
Sister	50	63	--	25
Grandfather	--	--	78	--
Grandmother	--	--	77	--
Grandson	--	--	33	--
Granddaughter	--	--	33	--

In another study of these same subsets of terms, D'Andrade (1965) performed a semantic differential by having each kin term rated on some thirty polar adjective scales. The two major factors were labeled "affect" and "boldness." The affect consisted primarily of a kind of desirable/undesirable dimension, and the boldness had to do primarily with activity.

Figure 6.

	m	f
+2	GrFa	GrMo
-2	GrSo	GrDa
+1	Fa	Mo
-1	So	Da
0	Br	Si

Jack Nadler performed an analysis of variance and fitted values by the method of maximum likelihood. Three dimensions emerged from this analysis -- sex, relative generation, and generation removed. The best model for his fitted values is shown in Table 5. By comparing the dimensions in Table 5 and those isolated previously in Table 4, it may be seen that the dimensions are isomorphic except that the factor analysis data reveals one additional distinction, namely, relative generation. Both figures also correspond to the data elicited utilizing the triad method as shown in Figure 5.

Table 5. Factor Loadings on "Affect" and "Boldness" for Lineal
Relatives and Fitted Values (Data from Roy D'Andrade
and Jack Nadler).

Affect

		Observed				Fitted Values	
		M	**F**			**M**	**F**
2	+	8.17	10.11	2	+	8.44	10.00
	-	7.29	8.60		-	7.08	8.64
1	+	9.15	10.92	1	+	9.19	10.75
	-	7.67	9.39		-	7.83	9.39
0		7.77	8.86	0		7.54	9.10

mean = 8.80   sex (f+) = .78   R.G. = .68

$G^0 = -.48$ $G^1 = .49$ $G^2 + -.25$

Boldness

		Observed				Fitted Values	
		M	**F**			**M**	**F**
2	+	8.12	4.17	2	+	8.06	4.59
	-	4.24	1.99		-	4.64	1.18
1	+	9.83	5.83	1	+	9.38	5.92
	-	5.61	2.50		-	5.97	2.50
0		8.14	4.15	0		7.88	4.41

mean = 5.45   sex (f-) = 1.74   R.G. = 1.71

$G^0 = .69$ $G^1 = .49$ $G^2 = -.84$

BIBLIOGRAPHY

Conklin, Harold C.
    1962   Comment on Frake, Charles O. The Ethnographic
            Study of Cognitive Systems. In Anthropology
            and Human Behavior, The Anthropological Society
            of Washington, Washington, D.C.

Berlin, Brent and A. Kimball Romney
    1964   Descriptive Semantics of Tzeltal Numeral Classi-
            fiers. In AA, Vol. 66, Number 3, June 1964.

D'Andrade, Roy G.
    1965   Trait Psychology and Componential Analysis.
            In AA, Vol. 67, Number 5, October 1965.

Guttman, L.
    1954   The Principal Components of Scalable Attitudes
            in Lazarsfeld, P.F. (Ed.), Mathematical Thinking
            in the Social Sciences. Glencoe, Ill.: The Free
            Press.

Lounsbury, Floyd G.
    1964   The Structural Analysis of Kinship Semantics.
            In Lunt, Horace G. (Ed.), Proceedings of the
            Ninth International Congress of Linguists.
            The Hague, Mouton and Co.

Ohman, S.
    1953   Theories of the "Linguistic Field" in Word, ix,
            123-34.

Romney, A. Kimball and Roy G. D'Andrade
    1964   Cognitive Aspects of English Kin Terms. In AA,
            Vol. 66, Number 3, June 1964.

Torgerson, Warren S.
    1958   Theory and Methods of Scaling. New York:
            John Wiley & Sons.

DISCUSSION

BAR-HILLEL: Am I right that you use semantic domain in a somewhat more liberal way than a logician would use a family of predicates?

ROMNEY: Yes; and the other thing in the discovery of the boundary is an empirical problem that is best done by doing some psychological type testing of where the natural boundaries are, but it is a relative thing, because if you put the two things in different context it affects their distance from each other; that is, if you get outside the domain. This will work if you are at that lowest level of contrast in a coherent domain.

SIMMONS: Have you any particular way of identifying domains? Presumably there are thousands of these falling across language.

ROMNEY: Well, that is a big empirical problem, and the thing is, for example, if you take Thorndike's Word List, listing role names in English, we've got something like 1200 terms. You have to start compartmentalizing that. We have used tests, judged similarity of various kinds. You use crude methods for your fist blockouts and then you use subtler and subtler methods and there are various techniques. I don't mean to make it sound easy.

For these personality trait terms we chose fifty. If they don't belong in the domain, the moment you put this measure on they really pop out. But to get them inclusive is very rough. There are some 5,000 registered color names in English. It's just fantastic!

BAR-HILLEL: What about sub-domains, refinements of domains, and things like that?

ROMNEY: Well, this is important and that's exactly what we
need work on. I have fuller data on each of these that I
didn't put in because I wanted to illustrate the method;
that is, other relatives and other emotional terms and other
color definitions.

ULLMANN: Do you distinguish between technical and non-
technical nomenclature?

ROMNEY: Yes.

ULLMANN: Some people would completely exclude scientific
terminology.

ROMNEY: That is what made me skeptical of taxonomy. I
haven't seen any folk taxonomy that has the properties
taxonomy should have. The only ones I know about are the
ones in science. They are probably real for the scientist,
but that will have to be tested.

GARVIN: I was just going to say that I am very pleased to
have this paper at this conference, because it introduces the
perspective that you get from looking at semantics through
culture, which was at one time the only way one was allowed
to do this in American linguistics. I don't think that just
because we now have greater freedom of choice we should ignore
this older way of looking at it.

   For instance, one of the things that Kim (Romney) probably
would have said, or could say, is that the problem of domains can
be handled by the observational and other techniques of the
cultural anthropologist. These things do pop out if you look
at things that happen in a society rather than merely just
think about what one ought to be if one did it, and that kind of
thing.

I think there is perhaps a little bit less introspection
in the cultural anthropologist's approach than there is in
that of the semantic theorizer, and this is to me very palatable
because of my particular personality.

SEMANTIC CLASSES AND SEMANTIC MESSAGE FORMS

Karen Sparck Jones
Cambridge Language Research Unit
England

Introduction

The paper which follows suggests an experimental approach
to semantic analysis. The semantic analysis of text presents
appalling, and indeed possibly insurmountable, difficulties;
but it is my belief that our ignorance is such that something
of value may be learnt even from quite limited experiments.
It may be that automatic semantic analysis of natural language
text is unattainable; but I nevertheless want to learn some-
thing, and though my particular rock pile may be a small one,
I shall dig away at it all the same. What follows is also
very simplified and schematic, since it is primarily intended
as a summary of my ideas for investigating one aspect of
semantic analysis, and not as a full-scale discussion of the
problem of semantic analysis as a whole.

****

One object of semantic analysis is to select the correct
meanings of words in text by using information supplied by
the surrounding linguistic context: basically, we have a
dictionary entry listing the possible meanings of a word,
and the features of the context which are required to specify
each one; and we have rules which define the procedure for
searching for these features. Now it is obvious that we
cannot operate a selection procedure which relies on really
detailed information about the meanings of words, or on the
occurrences of specific words: that is to say, we cannot
have a procedure such that, for example, we select sense

1 of word "a" if some other word in the surrounding context
has the meaning: is a squiggly kind of wirecutter for
manufacturing bone buttons, or if the context contains the
specific word "pliers." Or, to take another example, if we
have the sentence "My aunt was chewing rock," we cannot
rely on the occurrence of the particular word "chewing" to
resolve the ambiguity of "rock," which in English may mean
candy or stone. It has long been recognized that some
simplification is needed, and that this may be achieved by
using a semantic classification: the argument is that the
language-user identifies the general concepts with which a
piece of discourse is concerned, and relies on these to
sort out the senses of the words in the text. We therefore
provide dictionary entries which note the general concepts
conveyed by words, and search the surrounding context for
any word classified by such and such a general heading:
thus in our first example we select sense 1 of "a" if the
concept 'tool' is suggested by another (unspecified) word
in the text. Again, we select "rock" meaning candy in the
second case, because we have the general concepts of 'eating'
and 'food,' and we know that these concepts of 'eating' and
'food' may go together in this kind of way, while the concepts
of 'eating' and 'stone' do not generally go together in this
way.

Semantic analysis, insofar as this is possible from
dictionary and text without external references, thus depends
on some indication of the general concepts which may be
conveyed by a word, that is to say, on a specification of the
semantic classes to which it belongs, and of the semantic
relations which may hold between it and other words in text;
and research in automatic language analysis must therefore be
concerned with the nature of a semantic classification or
thesaurus, and with the nature of a semantic message unit.
With some understanding of these, we can then proceed, for

a given language, to the construction of a vocabulary classi-
fication and a listing, in terms of these classes, of accepted
message forms. There is indeed a third side to textual analysis,
namely that of matching an actual piece of text against the
list of message forms, to see which one of the set of permis-
sible message forms actually fits the text and can therefore
be selected to resolve the ambiguity of the particular words
in the text. This matching of text against dictionary and
inventory, however, depends on the prior existence of both
dictionary and inventory, and I shall therefore disregard it
here in order to concentrate on the actual construction of the
classification and obtaining of the inventory.

The two questions we are concerned with thus are: 1) what
is a semantic class? and 2) what is a semantic message form?
In the first case we have to take account of the paradigmatic
relations between the words in a vocabulary, and in the second
we have to consider the syntagmatic relation between the words
in a text. In the first case we have to say what it is for
a word to convey a concept, and in the second what it is for
concepts to go together; and we then have to say which concepts
are conveyed by which words and which concepts go together.

The first of these two questions has received more
attention, at least in the sense that a large variety of
semantic classifications have been constructed for different
purposes; the second remains obscure. This is to some extent
due to the fact that many classifications have been set up for
purposes like information retrieval, where there is no direct
application to text analysis. In discourse analysis, on the
other hand, we must take the classification and the way it is
used together. The connection is clearly shown, for example,
in Katz and Fodor's discussion of semantic analysis, and it
has been studied, for example, by Weinreich and members of
the Cambridge Language Research Unit. My object here is to

consider briefly what a semantic class might be like, what
a message form or type might be like, and how the use of a
particular kind of classification may influence the description
and treatment of message types, with a view to throwing some
light on all three questions.

A very simple model:

A word is a member of a semantic class if it expresses
the idea for which the class label stands; the members of a
class will thus be more or less synonymous, or at least close
in meaning, when compared with the rest of the vocabulary.
Thus, "run," "bound," and "spring" may appear in a class
labelled MOTION or ACTION. And if a word has several
meanings, it will appear in the appropriate different classes.

We now consider message types of the following form:
we have a topic and a comment, where we give the topic item P
and say that it has a P character. Thus given the sentence
"The professor was lecturing," we have a topic and a comment
which both come under the general heading TEACHING.

The general rule for selecting the correct use of ambiguous
words, and so effecting a semantic analysis of a text, is as
follows: if a piece of text is assumed to be semantically
repetitive, take the semantic class lists for the words in it,
look for recurring classes, and select as correct those meanings
of the words in the text which are defined by the recurring
headings.

This model is obviously appallingly naive: it will clearly
not work for a sentence like "The hippopotamus was feeding."
It nevertheless does work sometimes: in some early and very
tentative experiments at the C.L.R.U., an effective resolution
of ambiguity was achieved. It can also be argued that the
model is not wrong so much as inadequate: it is not the case
that semantic analysis can never be effected by this procedure,

but that it can only be carried out if the text concerned is platitudinous enough. What, therefore, should we do when our texts are more interesting and informative?

Katz and Fodor, though they are essentially concerned with this problem, do not put forward any very concrete suggestions. In the simple model just described there is no distinction in a dictionary entry between the semantic headings used to describe the meanings of the word and those used in analysis: in analysis we look for repetitions in the lists of headings which specify the meanings. In Katz and Fodor's standard entries there is a division between the headings which describe the word and those for which a search is made in the entries for other words in the text: thus one sense of "colourful" is defined by the heading (Colour), and has attached to it a note that this sense is selected if some other word in the text is describe by the heading (Physical Object). Analysis on this basis is thus more sophisticated than analysis by the repetition model, since to select the sense of colour- ful we require not that some other word should also be classi- fied by (Colour), but that some other word should be classified by (Physical Object): our analysis procedure is not confined to dull texts about colours being coloured, but can be applied to more interesting texts about things being coloured. And Katz and Fodor's assertion that textual analysis depends on the semantic relations which hold between the words in a text can be illustrated by our saying that there is a semantic relation between thing words and colour words. In the simple model we have only a trivial semantic relation, which we may crudely call the identity relation. If we use the terminology of message forms, we can say that Katz and Fodor are making use, in their entry for "coloured," of the fact that a great many individual messages can be generally described as being concerned with the physical properties of objects. Thus if we formulate

our message types in a very simple way for illustrative purposes, with semantic classes indicated by the letters A, B, etc., we have message types in our simple model of the form A,A, or A IS A, while in Katz and Fodor's model we have ones of the form A,B, or A IS B.

The work being done by Margaret Masterman and others, especially Wilks, at the C.L.R.U. approaches the same problem from a rather different angle: here the information about the way in which a word participates in messages in not attached to its entry; the meanings of a word are defined by semantic classes, as in Katz and Fodor, but the analysis of a text is performed by tests which see whether the words in it, when particular class specifications are selected for them, will fit into the 'slots' in some member of a given list of general message types. Thus, to give a greatly simplified example, if we have a list of message types including THING HAVE COLOUR and MAN HAVE ATTITUDE, and have the sentence "Flowers are red," we may select the correct sense of "red" and not the meaning 'socialist' because with "flowers" classified by THING and "red" by COLOUR, we can fit the sentence to the first message type, but not to the second, and with "red" classified by ATTITUDE and "flowers" by THING we cannot fit either. It must be emphasized that this is a very crude summary: the detailed approach is more sophisticated. For my purpose it is, however, sufficient: the important point is that though we have just considered two different suggestions as to how the problem of semantic analysis is to be tackled, they do share the same important feature, namely, that they depend on the existence of a list of message types: in one case the information represented by this list is largely incorporated in the dictionary entries, though some of it is incorporated in the rules for proceeding through the sentence in analysis, while in the other the list is used as it stands; but this difference is not important in this context.

The main criticism which can be levelled against both of these approaches is that they do not show how message types are to be set up, or give any criteria for judging whether given types are correct. It is of course unreasonable to demand definitive rules or criteria, but more discussion of these questions is required than is given. In this respect Katz and Fodor's deficiencies are much more glaring: the few scraps of information that can be gleaned from their very minimal examples are so small as to lead one to suspect that they have not really faced up to the problem at all. In Masterman we find much more substantial examples, and ones which are convincing as they stand. They nevertheless suffer from the major defect that they have been constructed with a particular list of semantic classes, and there is no reason to think that this list is better than another: it was indeed obtained a priori and might therefore be worse than some others.

The fact that these message types are formulated in terms of a particular set of classes, however, merely emphasises the point that the classification system and list of message forms required for semantic analysis are necessarily interdependent so that a particular choice in one case will influence the choice in the other. At the same time, the fact that message forms cannot be given except in terms of classes, though the reverse does not hold, suggests that we should start by attempting to set up a semantic classification, though this may be modi- fied as a result of our subsequent experience with our inventory of message forms. What we have to try to estimate, therefore, is what the effects of different kinds of classification will be. In my discussion I shall treat a semantic classification as a thesaurus, but it must be emphasized that my remarks about thesaurus headings or semantic classes apply equally to semantic components or semantic markers, and so on: essentially these are different names for the same thing, and in this context it does not really matter which we use.

I have discussed the various forms which a thesaurus
may take in detail elsewhere, so I shall simply say here
that in general a thesaurus class consists either 1) of a
set of words which are more or less synonymous or similar
in meaning; or 2) of a set of words which stand for objects
having a common property, such as being a receptacle; or
3) of a set of words which are characteristic of a particular
subject field, like agriculture, but which are neither similar
in meaning, nor represent objects having a common property
in any very significant sense.

In general, the approach to classification adopted by
both Katz and Fodor and by the C.L.R.U. can be described in
the terminology of the simple semantic model put forward
earlier, that is, by saying that a word is assigned to a class
if it conveys the general concept represented by the class
label: Katz and Fodor, for example, talk about resolving the
meaning of a word into its constituent atomic concepts.
Unfortunately, this kind of description of a class is so
vague that it can apply equally to the three types of class
I have distinguished. I have tried elsewhere to pin down
this notion of a semantic class, so that words may be subsumed
under headings in a reliable way; but this is not very easy,
and the whole approach suffers from the serious defect that
the list of headings is essentially a priori, though of course
it may be modified in practice in the course of classification.
The basic problem about constructing a semantic classification
is indeed that we get involved in every kind of difficulty
if we try to set up a list of semantic headings and then attempt
to sort words under them by asking, for each word and each
heading, the question "Can this word convey this idea?" I
have therefore argued that an alternative approacn should be
adopted in which classes are built up on some quite different
basis, and in particular have suggested that they may be
obtained from initially very small sets of words with synonymous

uses, by grouping sets which share common words: the results of
this application of the theory of clumps will then be classes of
synonyms and near-synonyms, that is, thesaurus classes of type (1);
and it is clearly a consequence of the method by which the classes
are obtained that the members of any class convey the same
general idea.

Given such classes, what effect will they have on the
description of message types? We no longer have a priori
classes, but we still have the problem of setting up message
types which we represent as strings of class labels, with
them. These types may have a more or less complex structure:
the simplest would consist simply of concatenations of labels,
while more elaborate ones woulu ' e some syntactic structure.
In Katz and Fodor this structure is not usually given explicitly
in the semantic dictionary entries, but is assumed to be
dependent on the detailed syntactic structure of the text,
which governs the course of the semantic analysis. In Masterman's
approach, on the other hand, the members of the list of message
forms have a simple syntactic structure determined by two
connectives and brackets. Unfortunately, while it seems reason-
able that message forms should have some structure, it is not
clear that either of these approaches is the correct one: Katz
and Fodor do not justify their assumption that the semantic
analysis of a text depends on its detailed syntactic description,
and in my view, the complexity of such descriptions is a good
reason for thinking that it does not. Masterman's method
represents an attempt to avoid just this difficulty, but is it-
self open to the objection that the syntax she adopts for her
message forms, and the particular message forms which she
gives in terms of it, are arbitrary.

How then are we to obtain our message types? One way of
doing it is to ask oneself what kind of things one says, and
to put these in a very general form. This is, in fact, what
Katz and Fodor and Masterman and Wilks are doing; and the

lurking danger is that we shall fall into the philosophers'
bog of predictableness, or start talking about objects and
properties: thus we set up the message type THING HAVE COLOUR
after concluding that leaves and books and ships may be
coloured and then find ourselves bothered by things like
windows, finishing in a morass of argument about shape
necessarily implying colour and colourlessness really being
logically the same as colouredness.

I want to put forward, not a solution to this problem,
since to attempt this would be reckless in the extreme, but
a suggestion as to how we may investigate what we mean by
a message type.

To do this, we must return to the naive model of analysis
described earlier. In this model text is treated as highly
platitudinous, since we look for conceptual repetition to
resolve ambiguity; and the defect of the approach is that
text is not so platitudinous. It is, however, arguable tha:
it is fairly platitudinous: much of what we say has to conform
to accepted general message types, or we will not be understood.
In this sense, though, we have a different sense of "plati-
tudinous": a piece of text is platitudinous because we have
heard the same kind of thing before, and not simply because
it is repetitive. I nevertheless wish to suggest that we
may be able to study the standard but not repetitive message
forms by proceeding from the simple repetitive form in a
controlled way. That is to say, I wish to try to throw some
light on comparatively informative message types of the form
A IS B by starting from the repetitive type A IS A.

To do this I shall refer to some of the consequences of the
method of obtaining semantic classes referred to above. This
method depends on the existence of small sets of synonymous word-
uses, or 'rows': a row, that is, contains the information t..at the
words whose signs appear in it are synonymous in one sense.

Moreover, the fact that the sign for the same word may appear in
several rows constitutes a semantic link between them: if two
rows share a high proportion of signs, we may infer that they are
semantically close; and it clearly follows that we can establish
'chains' of rows, linked by common words, where the length
of the chain indicates how close the words, whose uses are
defined by the end rows, are semantically. The detailed
consequences of the use of rows, and the various semantic
relations that can be interpreted with them are discussed
elsewhere: the important point is that the vague notion of
semantic distance can be pinned down, and that we can make
precise measurements of different degrees of semantic like-
ness.

The definition of a semantic class as a set of rows
with strong mutual overlaps in terms of common words is a
natural development from the starting point given by these
connections between individual rows; and as we saw earlier
such classes will consist of synonyms and near-synonyms, or
words which are close in meaning.

Now if we use classes of this kind in analysing text,
in the repetition model we will be looking for the same class;
but we do not want to confine ourselves to this, but want to
be able to use different but related classes. There must be
some relation between the classes constituting a message type,
by definition: the problem is to identify the semantic relations
which link classes in an acceptable or sensible message form.
However, though we know that this is a considerable problem,
we have been able to define some relations, namely those which
come under the general heading of relations indicating likeness
or similarity. These play their part in determining classes;
but they also hold between rows and words which are not members
of the same class, because far more quite distinct rows will
be chained to a given row, especially by long chains, than can

be accomodated in classes depending on heavy overlap in terms
of common words. And if these linked rows are not members of
the same class, then they will naturally be members of different
classes, given that each row for a vocabulary is a member of
some class.

From this point we can proceed as follows. We wish to
extend our range of messages from repetitive ones like A IS A
to non-repetitive ones like A IS B: and a natural way of doing
this is to consider first message types which say not that
A IS B, but that A IS A'; that deal not with concepts which
are quite different or distinct, but with ones which, though
they are not the same, are like one another. As we have seen,
the assumption on which any approach to semantic analysis must
be based in that discourse has some degree of semantic or
conceptual coherence, that it deals with concepts which go
together. The sense in which two identical concepts go
together is a trivial one; and the sense in which two quite
distinct concepts go together, on the other hand, is just what
we have difficulty in pinning down. The sense in which two
like or similar concepts go together, however, is neither
wholly trivial nor impossible to define. We must of course
eventually deal with quite different or contrasting concepts
which go together, but in the absence of any very clear idea
of what it is for two concepts to go together, we can, I claim,
justify the attempt to walk before we can run. Suppose, there-
fore, that we concentrate on message types dealing with similar
or close concepts. My argument is that we can obtain such
message types by considering semantic classes which are linked
in the way I have just described: that is to say, if we have
two classes which are different but are linked through common
rows or chains or rows, then they are prima facie candidates
for message types of the form A IS A', and possibly also for
those of the form A IS B.

This in itself, however, is not enough, since the absence
of any restriction on the length of the chains may give
connections between virtually any pair of classes. Moreover,
the assertion that any pair of linked classes are candidates
for a message type does not do much to help us in actually
identifying particular pairs; there are far too many possible
chains for us to explore them all.

But we may nevertheless obtain pairs of classes which
are not too tenuously linked, and in a comparatively much
less exhausting way. To do this, we make use of higher level
classes, that is, classes of our initial classes. If these
are obtained by analogous methods to out initial classes, they
must consist of linked classes, and moreover of classes which
are fairly strongly and mutually linked. We will thus obtain
specific pairs of classes, representing pairs of concepts,
which are semantically quite close, and without the appalling
effort of finding whether every pair of our initial set of
classes is linked by some chain. The pairs of classes which
are members of a specific second-order class can then be
combined in simple message forms, and these can be used
experimentally as a basis for further investigations of the
way in which analysis may be performed. Thus if we have a
higher-level class containing the classes P, Q, R, S, we will
have PQ, PR, PS, QR, QS, and RS, as accepted message forms.

The foregoing argument may be illustrated as follows:
experiments so far carried out suggest that the automatically
obtained groups clumps, of rows which we set up initially will
be quite like the sections in Roget's Thesaurus which consists
largely of synonyms and near-synonyms, like Number 682, Activity,
which contains words like "briskness," "liveliness," "agility,"
"smartness," "quickness," "speed," "movement," "bustle," "hustle,"
"hasten," "brisk," "lively," "alert." (It is difficult to be
more precise since experiments so far have not been on a very
large scale.) In the Thesaurus there are cross-references
from this section to others like 282 Progression and 686

Exertion, and as these are defined by common words, we can
infer that any similar groups of rows would be strongly
linked, and would therefore be grouped together by a second
round of classification. And we then find that we have
concepts which go together, but are not the same, like
'Work' and 'Progress.'

On this basis, what do message forms look like? The
fact that we have two different concepts which go together
suggests that we can tackle sentences like "The work is
progressing" or "The labour is advancing": the question is
what form should our message types take, given that we know
which class labels we should combine in them, and how should
they be used?

The simplest approach would be to take simple concatenations
of classes, without any structure, as message types: they would,
after all, indicate permitted combinations of ideas, and this
is what the most minimal message type is. Thus given P and Q,
PQ is the same as QP. The procedure for identifying the
message type underlying a text would then be a very elementary
one, representing a combination of that used in the simple
model outlined earlier and that used by Masterman: given the
class lists for the words in a text, we would see whether
any particular selection of classes for the words would match
our list of permitted combinations. We would thus be fitting
our words into slots as in Masterman, but would not be using
ordered slots, as in the simple model. Thus to refer to our
example, we would have a list of message forms including
Activity, Progress; then, given the sentence "The work progresses,"
we would find that the class membership lists defining the
different senses of "work" and "progress" include Activity and
Progress; and since inspection of our list of permitted combi-
nations shows that these two may go together, we select the
corresponding senses of "work" and "progresses," and eliminate,

for instance, "work" meaning froth. Of course this example
is grossly oversimplified: I am concerned primarily with
indicating how we might set up message types which do not
simply represent conceptual repetition, and how we might
use them to analyse sentences which are not wholly plati-
tudinous.

However, since the concatenating message type is certainly
too simple, in the way in which the repetition model was too
simple, we should consider how we might proceed to more
sophisticated, that is structured, message forms. As noted
earlier, Masterman's message forms have a syntactic structure,
so that, for instance, we have MAN HAVE STUFF, the use of
syntax being a device to exclude the unwanted interpretations
which may be derived from unstructured types, such as STUFF
HAVE MAN, and to take note of structure which exists in the
text which is being analysed. Again, Katz and Fodor's message
types are structured. We do not, however, want to have to
think up possible structures, or much of the effort we have
gone to achieve objectivity will be wasted: we have obtained
our semantic classes objectively, and though this is clearly
a gain which we will retain, it would be nice not to be forced
to set up our structures a priori, but to construct them in
some less subjective way.

One possible approach to this would be to take actual
sentences, and to substitute classes for the words in them
while preserving the sentential syntax. The important point
is that these would not be any sentences, since the substi-
tution would then be open to the criticism that it represented
a vicious circle, but would only be tautologous or analytic
sentences, on some intuitive interpretation of analyticity.
On this basis we would make use of sentences like "The singer
sings," and adopting a message form of the 'topic-comment'
kind, would permit combinations in this form of all the classes
which are grouped in a higher level class with that in which

this sense of "singer" occurs, and those which are grouped
with that in which "sings" occurs. Thus we might, again
referring to Roget for an example, obtain a permitted combi-
nation of this kind containing a group like Roget's section
524 Interpreter and one like 582 Speech, so that we could
attempt to analyse sentences like "The diplomat argued."
The extension of this approach using longer forms derived
from considering sentences like "The singer sang a song"
would then clearly be possible. Of course we rely in doing
this on some language-user's assertion that a sentence like
"The singer sings" is analytic in some intuitive sense, and
perhaps also that some less obviously tautologous sentences
like "The spinster is unmarried" are so too; but in this we
are relying on the language-user's knowledge of his language
in the same way as we rely on it to construct rows, that is
assert that two words may be substituted in a sentence. It
is arguable that any lexicographic work depends on someone's
knowledge of the language somewhere, if only to gauge the
significance of results which have been obtained mechanically;
at the same time we want to damp down the possible uncertainty
that this involves: and this would be one way of introducing
structured message forms which, while accepting the need for
reference to a language-user, prevents him from making too
many idiosyncratic responses.

This approach is naturally a tentative one; but it
nevertheless represents a concrete suggestion as to how the
problem of setting up a list of message forms might be tackled.
It does not deal with the question of how an actual piece of
text, with its detailed structure, is to be matched against
a much more summary structure; but this arises is some form,
however we obtain our message types, and may therefore be
considered separately. We may in any case learn much about
linguistic analysis by starting only with very simple pieces
of text where this matching problem is minimised.

References

1. Sparck Jones, K., _Synonymy and Semantic Classification_,
   Ph.D. Thesis, University of Cambridge, 1964; Cambridge
   Language Research Unit, mimeo, 1964.

2. Katz, J.J. and Fodor, J.A., "The Structure of a Semantic
   Theory," _Language_, Vol. 39, 1963, p. 170.

3. Masterman, M. _et al._, "Semantic Basis of Communication,"
   Cambridge Language Research Unit, mimeo, 1964.

4. Wilks, Y., "Computable Semantic Derivations," Cambridge
   Language Research Unit, mimeo, 1965.

DISCUSSION

ROSS: What happens, for example in "The singer sings.",
when you have super classes like this, taking a simple
sentence, for example, a subject-predicate sentence? I
don't understand how you can disambiguate in a case like
this.

SPARCK JONES: Yes. I possibly was not clear enough about
this. This is not a sentence I am now trying to resolve
the ambiguity of. I am using sentences like this as devices.
I am assuming that I myself can resolve the ambiguity. I
am using them as devices for obtaining structured message
forms. If you use super clumps, all you get is permitted
combinations of concepts, and combinations of concepts are
not going to do enough for you. The fact that "eating" and
"food" go together is not enough in a case like "My aunt was
chewing candy" because it's the food that is eaten, it is
not something that is eaten by the food.

 I had available some super clumps or super clusters.
I was trying to use sentences like this to obtain structured
things like A-A', so that the other concepts which share
this super class were the ones that define this use of "singer"
and the other super class were the ones that define the use
of "sings"; that they can all be put together in some kind of
structure like that.

VON GLASERSFELD: I think what you said about the necessity
of indicating the function between your classes -- in other
words, when you have "food" and "eating" -- it is not enough
to say they belong to the same field, but that "food is in
a certain relation to any activity that can be called "eating"
is very important. Ceccato's group, some seven or eight years
ago, worked on that very seriously and tried to establish

what they called the notional sphere, which is precisely
that kind of intuitively arrived at classification with
the indication of the relations between the classes.

This became so complicated that it was very difficult
to handle because it is extremely difficult to see where
these functions, the relations between the classes, should
stop. I think some way has to be found to determine which
functions are necessary for disambiguation and which not.

But there is another point that I think is important,
in the application of what you get out of these classifi-
cations. You talked at the beginning about "permitted
message forms." I think there is an criginal mistake in
that the sense that even your example shows, you can't
possibly exclude absolutely certain senses. I think the
only useful information you can draw from these classes and
classifications is probablistic. This is not a criticism of
yours; it goes against Katz and Fodor just as much. It goes
against anyone who wants to say "This is not allowed," because
if you have an example, "My aunt has had a relapse; last night
she was chewing the bed post," that is perfectly possible,
and bedposts are not to be eaten.

SPARCK JONES: This is the basic problem.

VON GLASERSFELD: It is probablistic. What you say is
extremely improbable; if she eats rocks, the "rock" meaning
of "candy" is very probable.

GARVIN: "Chewing" is also not merely "eating." You could
also have "the crusher chewing the rock."

MASTERMAN: There seem to be difficulties about this model,
but surely, as soon as you get into semantics the classes will
come together in a super class.

SPARCK JONES: Yes. I think the classes that come together
in a super class would represent some uses of the words
concerned, and not the others.

XI.

SEMANTIC SELF-ORGANIZATION

Eugene D. Pendergraft
Linguistics Research Center
The University of Texas

1. INTRODUCTION

This is a supplement to the paper on Automatic Linguistic Classification that Pendergraft and Dale [1] presented last May in New York to the 1965 International Conference on Computational Linguistics. Since then a somewhat fuller and up-to-date account of our experiments with syntactic self-organization has appeared in the form of a working paper [2]. My aim here is to indicate how we plan to extend our experimental design to include relations that may be characterized as semantic rather than syntactic.

Essentially the extension will involve consideration of the next level of the hierarchical linguistic model [3] we have been studying and the development of algorithms capable of self-organization at that higher level. Thus our next objective will be semantic self-organization within the tentative but specific frame of our formal working hypothesis.[4]

As in our earlier papers, automatic classification will be regarded as consisting of those operations that, when sussessful, result in a taxonomy of objects based on their empirically given properties or relations. Self-organization will imply additionally that there are operations evaluating the taxonomy and modifying it in such a manner that it should tend to improve.

In this view a self-organizing system is one carrying out a particular strategy in automatic classification, the strategy being especially suitable when the properties or relations of a large universe of objects are presented over a period of time in successive experiences. Each experience may contribute new evidence about the way the objects should be classified. Since knowledge of the objects may be incomplete at any stage of processing, the taxonomy can be expected to change dynamically in response to the accumulating evidence. But the strategy would work equally well with objects whose properties or relations were known at the outset. Although the same objects would then be presented repetitiously in the successive experiences, new evidence might be extracted from them on each presentation.

The key to the strategy, therefore, is a processing cycle in which deduction and induction alternate. From their empirically given properties or relations, the objects presered ed in each experience are deduced to be members of particular classes in the current taxonomy. Various statistics are then collected on relations between inferential events in this deductive process. By means of automatic classification, appropriate modifications in the taxonomy are induced from these statistics. Finally, the taxonomy of objects is updated in preparation for the next cycle.

This strategy is novel in that it applies automatic classification to what is being deduced about the objects presented in experience rather than directly to those objects. Accordingly, one must distinguish between automatic classification of the events of deductive inference about the objects and automatic classification of the objects themselves. In each processing cycle the former is a prerequisite to the latter. From the resultant classification of deductive events, the inductive operations will infer how certain classes of the objects may be specialized or generalized, or that two or

more of the classes may in fact have identical membership,
or that certain relations may exist between the classes.
The specific inductive operations used in our experiments
have been explained in detail $\underline{/}^-2\underline{\_/}$ and will only be
mentioned herein.

An advantage of the strategy is that it operates at
a higher level of abstraction than automatic classification
applied directly to individual objects. The problem of
dealing with a large universe of objects may be reduced to
a number of subproblems concerned with individual classes
or collections of classes. Another advantage is the
possibility of considering parts of the universe in succession.
More tractable processing requirements are a consequence of
these advantages.

Lastly, the strategy appears to be general in the sense
that it can be adapted to various universes of objects and
their properties or relations. The following paragraphs
discuss such an adaptation, whereby self-organization now
being applied to a taxonomy of lexical segments based on
their syntactic relations will be adapted to a taxonomy of
syntactical segments based on relations among them that have
been characterized as semantic. A justification of this
characterization will not be given; however, a few remarks
may be helpful in pointing out some semantical aspects of
the problem.

2. INDUCTION FROM PREDICATES TO CONCEPTS

Attempts to classify the predicates of relations in terms of the predicates of their arguments, and vice versa, have usually taken signs of the predicates to be lexical segments $/^-5,6,7\_7$. The ultimate aim of such experiments is a method by which to proceed inductively from representations of predicates in natural language to representations of concepts, and thus from statements to prepositions, for the purposes of automated information retrieval, translation or the like. It seems plausible that such heuristically derived classes of predicates might correlate with concepts, though as yet the results have not been convincing.

In recent years methods of mechanical translation have been developed in which syntactical rather than lexical units are substituted interlingually $/^-8,9\_7$. Whatever the formal assumptions underlying a system of this kind, parts of the syntactic taxonomy of one language must be equated to parts of the syntactic taxonomy of another. Presumably those corresponding parts will be the ones needed to recognize predicates in the first language and to produce equivalent predicates in the second.

Consequently the alternate possibility has emerged that the signs of predicates in natural language may be syntactical segments, that is, those parts of the syntactic taxonomy needed either to recognize or to produce the predicates. Much as lexical segments may be conceptualized as constructions of phonemes or graphemes, then, syntactical segments may be thought of as constructions of syntactic rules. The constitutive relations between objects of these two fundamental types would of course be different. For example, in our hypothesis these distinct relations are referred to as "concatenation" and "application" respectively $/^-4,10\_7$.

With appropriate constitutive relations between syntactical segments, induction from predicates to concepts may obviously be

approached as well by predicates represented syntactically as by predicates represented lexically. To study this possibility of semantic induction is the basic objective of the experiments in semantic self-orgainzation which we propose to undertake.

3. SPECIFICATION OF SEMANTIC STATISTICS

For our experiments with syntactic self-organization
a system of computer programs has been developed $/^-2_7$
primarily combining the deductive capability of automatic
syntactic analysis with the inductive capability of auto-
matic classification. Provision has also been made for
storing the syntactic taxonomy, for accessing it as a basis
for automatic syntactic analysis of texts, for collecting
and storing the statistics on inferential events in the
resultant syntactic analysis, for accessing the syntactical
statistics as a basic for automatic classification, and
for modifying the syntactic taxonomy in the ways induced
from the statistics.

The texts presented for automatic syntactic analysis
therefore constitute the experiences of the system. Each
text may be of any length or may transcribe any language.
Nevertheless, to limit the volume of statistical data
collected from analysis results, we have found it profitable
in our experiments to use texts of about 2000 running words
for each cycle of deduction and induction. These are being
taken from the Brown University corpus of one million running
words of contemporary English $/^-11_7$.

Our present approach to automatic morphological classi-
fication, that is to say to the problem of what lexical
segments should be classified syntactically $/^-2_7$, is to
merely perform graphemic analysis on the texts as a prelude
to syntactic analysis. From the syntactic taxonomy of
graphemes, we will then try to extract morphemic segments by
means of entropy computations. Automatic semiological
classification will be attacked by the analogous approach
of performing tagmemic analysis as a prelude to semantic
analysis and extracting sememic segments from the resultant
semantic taxonomy of individual syntactic rules $/^-2_7$.

Programs for the deductive phase of the semantic cycle
have been completed. What must now be specified are the
semantical statistics which will be used by programs in the
inductive phase of the cycle.

Four types of syntactical statistics are being collected
and used in the current programs:

Type 1: Rule Use

The frequency of use in automatic syntactic analysis of
each syntactic rule is recorded as a basis for automating
assignment of syntactic rule probabilities. The class name
in the left side of the rule will also be recorded.

Type 2: Rule Application

The frequency of application of the syntactic rule Y at
position p in the rule X (i.e. the frequency of the event
$X^p Y$) is recorded for the pair (X^p, Y). These are the
incidence data for the automatic classification operation
which specializes syntactic classes. Thus it is necessary
to distinguish (by means of a descriptor in a statistical
store) the particular syntactic class which is symbolized
at position p in the rule X. Only those statistics in the
substore so distinguished (by that descriptor) are needed
in the specialization operation which subdivides that class.

Here "position p" refers to the p-th variable (specifi-
cally neglecting the constants) in the right-hand side of
the rule, rather than to the superscript associated with that
variable. The two naming schemes may sometimes be identical,
because automatically generated rules will have the super-
scripts numbered consecutively from left to right. This
consecutive ordering is only tentative, however; superscripts
ordered differently are required in the semantical classi-
fication operations. Consequently, the naming scheme utilizing
the actual symbol positions in syntactic rules will be
employed not only in these programs but in the programs that
modify the syntactic taxonomy.

Type 3: Class Coincidence

The frequency with which any lexical segment is analyzed ambiguously as a member of both the syntactic class A and the class B is recorded for the pair (A,B). The operation of class identification is based on these (symmetrical) incidence data. Class generalization, the operation sometimes performed as an alternative to class identification, is based on incidence data which are assembled automatically from the results of the class identification operation.

The term "lexical segment" refers here to any uninterrupted sequence of characters representing either graphemic or phonemic inputs. The segment has a "beginning" and an "end." During processing the beginning of the segment is named by the character position preceding it and the end by its own character position. Character positions are numbered con- secutively through the entire input sequence.

Type 4: Class Concatenation

The frequency with which any lexical segment in the syntactic class A (as determined by automatic syntactic analysis) is concatenated to one in the class B (i.e. the frequency of the event A⌢B) is recorded for the pair (A,B). A distinction is made (by means of a descriptor) between those pairs separated by a blank character and those which are not. The two (the "blank" and "non-blank") sets of incidence data are processed independently as inputs to the operation which generates new syntactic rules.

All programs which collect syntactical statistics and update the statistical stores have been completed and are in use. Programs have also been written to remove from the stores any rule number of class name that no longer occurs in the syntactic descriptions.

Descriptors will be used in the statistical store to distinguish the semantical from the syntactical statistics.

Morphological and semiological statistics will be added
later, the final store having the order:

        (a) morphological

        (b) syntactical

        (c) semiological

        (d) semantical

Semantical statistics will be analogous to the syntactical
Exceptions in the four types are noted below:

Type 1: Rule Use

The frequency of use in automatic semantic analysis of
semantic rules will be recorded and employed in automating
the assignment of semantic rule probabilities, exactly as
in the syntactical case. The semantic class name in the
left side of the rule will be recorded also, as in the
syntactical statistics.

Type 2: Rule Application

The frequency of application of the semantic rule Y
at position p in the rule X (i.e. the event $X \overset{p}{} Y$) will be
recorded for the pair (X^p , Y). As in syntactical case,
the class symbolized in the rule X at position p will be
distinguished (by means of a descriptor) so that the appro-
priate subset of incidence data can be located to subdivide
that class. Again "position p" will refer to the p-th
variable in the right-hand side of the semantic rule, not
to the superscript associated with that variable. Since
the new classes resulting from the class specialization
operation will have the some degree as the one which was
subdivided, it will not be necessary to carry any informa-
tion about the degree of semantic classes in these statistics.

Type 3: Class Coincidence

The frequency with which any syntactical segment is
analyzed ambiguously as a member of both the semantic class A
and the class B will be recorded for the pair (A,B) The
operation of semantic class identification will be based on these
incidence data, and, as before, semantic class generalization

on the results of identification.

The inputs of automatic semantic analysis will represent the syntactic "trees" that resulted from automatic analysis of the lexical inputs. Hence "syntactical segment," as used above, refers to some part of a syntactic tree. That part, being itself formed as a tree, will have a "root" and one or more "branches." Or it will have a root and no branches, being in this case a "terminal" segment within the tree as a whole. Each syntactical segment will also have a "degree" determined by the number of its branches.

A syntactical segment is identified during processing by the position of its root and each of its branches. Because each syntactical segment subtends a definite lexical segment (i.e. that part of the lexical inputs which it analyzes), its root can be identified in part by the character position of the end of the lexical segment it subtends. The current naming, scheme, in addition, assigns a unique "entry" number to each root partially named by the same character. The branches of the syntactical segment, if there are any, are named according to the roots they adjoin in the overall treee. In particular, each branch joins a unique root and has the same name as that root.

According to the semantic hypothesis, all of the syntactical segments which are the members of a particular semantic class must have the same degree. The "degree" associated with that semantic class is (by definition) the same degree as each of its members. In consequence, it would be impossible for a syntactical segment to be analyzed ambiguously by two semantic classes with different degrees.

Since coincidence cannot possibly occur in the outputs of automatic semantic analysis between classes with different degrees, no test of this condition will be needed in programs which collect these statistics. But the operations of speciali-zation and generalization will be performed independently for each degree (to conserve space in automatic classification). As a consequence, the degree of each semantic class will be dis-tinguished (by a descriptor) in the statistical store.

The fact that the semantical metalanguage may have synonyms (i.e. one syntactical segment may have different names $\underline{/4\underline{/}}$) poses another technical requirement, both in these classification operations and in semantic rule genera- tion. Besides its degree, each semantic class will have an associated numeral called its "status." The status of any class of positive degree which has not been introduced as a result of automated rule generation will be the numeral one. If (as a result of automated rule generation) the semantic class γ is intorduced by means of the new rule $\gamma \supseteq \propto^{i} \beta$ then the status of γ will be the numeral i. The status of any class of degree zero will be zero.

By using this status information about semantic classes, it will be possible for the operation which generates semantic rules to limit each new construction to a standard form, <u>viz</u>., the superscripts at the successive points between syntactic segments classified by the construc- tion will be in nondecreasing order. The fundamental strategy will be to permit synonymous constructions, but to generate new constructions only in the standard form.

Each new automatically generated syntactic rule will be duplicated with its superscripts in the reverse order. The original rule and its duplicate will then be placed in a unique semantic class, as will each new syntactic rule which was not generated automatically but coded manually.

To prevent the identification of synonyms, not only the degree of each semantic class but its status will be distinguished (by separate descriptors) in the coincidence statistics. The operations of class identification and generalization will process the classes having a particular status independently, even though classes differing in status may have the same degree.

It will be necessary for the collection programs to test the status of each class before recording coincidence, to ensure that the coinciding classes have the same status. If they do not, the coincidence will not be recorded. (Note that the status of a semantic class may be greater than its degree.)

Type 4: Class Concatenation

The frequency with which any syntactical segment in the semantic class B (as determined by automatic semantic analysis) is joined to one in the class A at superscript p (i.e. the frequency of the event A^pB) will be recorded for the pair (A^p, B), provided p is not less than the status of A. If p is less than the status of A, the event will not be recorded.

Semantic class concatenation statistics will require the following distinctions (by separate descriptors):

(a) the degree of A
(b) the status of A
(c) the degree of B
(d) the superscript p

The degrees of A and B will be required by programs which actually encode the automatically generated semantic rules, as will the information about superscript p. The status of B need not be tested, as will be the case in collecting coincidence statistics.

The distinction in syntactic class concatenation between the pairs separated by a blank and not so separated will not be appropriate in the semantical statistics. Furthermore, the distinctions listed above will be made solely for the purpose of rule encoding, they will not demarcate subsets of incidence data to be processed independently. In the pairs of component sets resulting from automatic classification, different rules will be encoded for the classes differing either in degree or in superscript at the point of concatenation.

REFERENCES

1. E.D. Pendergraft and N. Dale, "Automatic Linguistic Classification," presented to the 1965 International Conference on Computational Linguistics, May 1965.

2. _____, "Automatic Linguistic Classification," LRC 65 WAT-1, Austin: Linguistics Research Center, November 1965.

3. W.P. Lehmann and E.D. Pendergraft, "Structural Models for Linguistic Automation," Vistas in Information Handling, Washington: Spartan Books, 1963.

4. E.D. Pendergraft, "Basic Methodology," Symposium on the Current Status of Research, LRC 63 SR-1, Austin: Linguistics Research Center, October 1963.

5. F.T. Sommers, "Semantic Structures and Automatic Listing of Linguistic Ambiguity," B-222, New York: Columbia University, October 1961.

6. K. Sparck-Jones, "Synonymy and Semantic Classification," Cambridge, England: Cambridge Language Research Unit, 1964.

7. B. Foreman, "An Experiment in Semantic Classification," LRC 65 WT-3, Austin: Linguistics Research Center, December 1965.

8. V.H. Yngve, "Implications of Mechanical Translation Research," Cambridge: Research Laboratory of Electronics, November 1963.

9. W. Tosh, Syntactic Translation, The Hague: Mouton & Co., 1965.

10. W.B. Estes, W.A. Holley and E.D. Pendergraft, "Formation and Transformation Structures," LRC 63 WTM-3, Austin: Linguistics Research Center, May 1963.

11. W.N. Francis, "A Standard Sample of Present-Day Edited American English, for Use with Digital Computers," Manual of Information, Providence: Brown University, 1964.

DISCUSSION

MASTERMAN: One of the things I very badly want to know is how you put in the initial probabilities which enable you, then, to see what is the most likely semantic output. I can quite see you can calculate probabilities once you put in probabilities. That's what probability calculus is for. I don't see, and you haven't said, on what grounds, on what sort of evidence, you put in your initial probabilities. Again I may be wrong about the relationship between the two systems.

Well now, from these clearly you get semantic classification, and I am personally very interested in these pairs and the binary relations. This is not fair, to ask you for anything except a reference, but if you have a paper on this particular point with an example, I will be grateful to have it.

PENDERGRAFT: All right. You have raised several questions. One is concerning your feeling about the experiments.

My feeling about the experiments is this, simply, that we are interested in the empirical result here, and I am not arguing for the experiment. It is an experiment already in progress, and the result is what we are interested in.

Regarding the formalization, the languages up to the pragmatic language, these formal languages were specified in our report in 1963 called Status of Current Research, Linguistic Research Center, under "Basic Metholology."

As for the issue of choosing between descriptions, we have all kinds of questions about adequacy of descriptions.

I should explain that one of the reasons we went into this in the first place is that we are engaged in translation experiments with English, German, Russian, Chinese, and about ten languages now. We have had almost seven years of experience

in writing these descriptions. We have noted, now that we have
large capacity analysis programs, that there is a great dif-
ference in the processing characteristics of grammars written
by different linguists, and this led us to suspect that there
is a property of grammars which hasn't been studied very well;
namely, those properties which would be concerned with how much
information is in the grammar. If you have a small description
with large categories, what generally happens in linguistic
analysis is that you get many results. You do too much process-
ing and you wind up with a lot of ambiguity. There just isn't
enough information in the grammar.

On the other hand, you know intuitively that you can have
too many distinctions, distinctions that are not necessary at
all. So what we are trying to specify here is what it means
to have just the right amount of information in the grammar.
In other words, we specify a procedure where the grammar starts
subdividing classes and does so until it reaches the place where
it hasn't any more formal information to further subdivide. So
we anticipate the convergence process here which would wind up
with an optimal grammar in the sense of a grammar having just
the right amount of information to make the distinctions we
are trying to make and no more.

This is not an experiment -- or I should say this is an
experiment to try to improve grammars as much as to discover
them. The programs have been written in such a way that we
can take an existing description and have the machine change
it. We don't have to start from absolute zero and try and have
the machine learn the language. These programs are intended
primarily to help our linguists in syntactic classification
and secondly in this semantic area, which we know even less
about. The machine will make suggestions to them in the sense
of writing rules, and they will be able to look at these and
see whether they agree with its prognostications.

I should say that we will have semantic translation
algorithms, programmed and finished by February, which is
just two months off. What we are concerned with is the very
great problems of getting together the data basis that you
will use in these very complex algorithms.

GARVIN: Is your term "application" the same as Shaumjan's?

PENDERGRAFT: I'll say just off hand "No."

GARVIN: It would be good to make it clear for the general
public because Shaumjan has become known to the general
public for his application in generative grammar.

ROSS: His use of the term is older.

PENDERGRAFT: As I said at the beginning, the key to the
whole business is that the constituted relation changes as
you go up the hierarchy. Our pragmatic language, you see,
comes off now in another direction, doing precisely the same
thing at the higher level.

XII.

A TAG LANGUAGE FOR SYNTACTIC AND SEMANTIC ANALYSIS

Warren J. Plath

International Business Machines Corporation
Thomas J. Watson Research Center

1. Introduction

This paper describes a problem-oriented language designed
for writing phrase structure parsing rules and briefly explores
some possibilities of employing the language in automatic
semantic analysis along line similar to those proposed by
Katz and Fodor[1]. In its current form, the language has
shown promise of serving as a powerful and convenient tool
for automatic syntactic analysis, owing largely to its
facilities for describing grammatical constitutents and their
relationships in terms of structured symbols, rather than
atomic ones. Although it appears that these same facilities
will also be of considerable value in semantic analysis of
the type considered here, even the simple example discussed
in the paper suggests the need for fundamental extensions of
the language, which appear to be motivated by syntactic consi-
derations as well.

The basic notational device employed within the language
to represent the substructure of constituents is the tag, or
attribute-value pair. A virtually unlimited number of tags
can be associated with any constituent name, whether it re-
presents a data item or appears as part of a grammar rule.
In the work carried out to date on automatic syntactic analysis
of Russian, using a parsing system based on the tag language,
the ability to introduce tags freely and define general opera-
tions on them has been instrumental in attaining such diverse
objectives as:

(1) efficient handling of grammatical similar subclasses;

(2) elimination of redundant multiple analyses; and

(3) effective treatment of agreement and government rela-
tionships involving grammatical attributes such as case,
number, and gender.

Preliminary indications are that, to the extent that semantic
attributes and their possible values can be defined, there is
little difficulty in expressing them in the form of tags. How-
ever, when it comes to writing rules describing selection re-
strictions involving multiple attributes (whether semantic or
syntactic in nature), the present form of the language turns
out to be considerably less convenient than one would like.
What appears to be required is an extension of the language
to include additional operations on strings of tags, analogous
to current operations on individual ones.

2. The Tag Language

Although there has been a tendency in much of the theore-
tical work on phrase structure grammar, as well as in experi-
mental work on automatic phrase structure parsing, to employ
atomic symbols in referring to grammatical constituents, the
systematic use of attribute-value tags in computational
linguistics goes back at least as far as the subscript notation
of Yngve's COMIT [2]. As is well known by those familiar with
COMIT, this rather general language for non-numerical proces-
sing has provisions for appending a virtually unlimited number
of logical subscripts to constituent names and includes special
operations for testing and merging the values assigned to the
subscripts. Indeed, the COMMIT subscripting system represents
one of two major influences on the present tag language, the
other being the system of grammatical indices developed in
the work on multiple-path predictive syntactic analysis of
Russian at Harvard[3]. The latter system is much more restricted

in scope than COMIT, since it was designed exclusively for writing predictive grammar rules involving a fixed inventory of grammatical attributes of importance in syntactic analysis. The chief innovation in this rather specialized language of grammatical indices is the employment of variables as index values, which makes it possible to write very general rules reflecting agreement and government relationships

The present tag lanugage shares with the grammatical index notation the property of being a rule-writing language in which variables play an important role, but it is also endowed with a COMIT-like facility for ad-lib introduction of names of constituents, attributes, and values. The language plays a central role in a parsing system known as the Combinatorial Syntactic Analyzer[+], which operates on the IBM 7094. If one temporarily disregards the overlay of tag operations, the parser can be described as using an exhaustive bottom-to-top analysis algorithm, currently limited to binary combination rules of the context-free type; that is, the rules are of the general form $C_1 + C_2 = C_3$, which signifies that whenever a constituent of type C_1 is immediately to the left of a constituent of type C_2, they can be combined to form a constitute of type C_3. The flow of the underlying algorithm, which is due to Kuno, differs from the Cocke-Robinson parsing logic[4] in that iteration is performed not on increasing constituent length, but by introduction of the word classes for the next word to the right whenever all combinations involving previous words have been attempted. However, since both algorithms eventually produce all combinations of adjacent constituents that are permissible with respect to a given grammar, they can be regarded as equivalent for purposes of the present discussion.

+ The original version of the analysis system was programmed by Robert Strom, who has also made significant contributions to the design of the tag language.

Within the parsing system there are two distinct types
of constituents: those that represent data items, i.e.,
particular instances of constituents in a given sentence
being processed, and those that appear as parts of grammar
rules. In the tag language, both types of constituents typically
consist of a part of speech or part of sentence name followed
by a (possibly null) string of tags. Each tag, in turn,
consists of an attribute name, a '/', and a list of one or
more value names. The following is an example of possible
coding for a data item-- the English word "shirt", described
as a concrete noun, singular number, denoting an object which
is neither animate nor human:

(1) NOUN SUBCLS/CONC NUMBER/SING ANIM/NO HUMAN/NO

COMIT users will note that, aside from a difference in punc-
tuation conventions and the fact that names are limited to
six, rather than twelve, characters, (1) is very similar to
a COMIT constituent with logical subscripts.

A more interesting example is (2), which illustrates the
way in which certain features of the tag language-- in particu-
lar, the employment of variable values of attributes--can be
used to advantage in writing grammar rules.

(2)   ADJ CASE/X NUM/Y GEN/Z ANIM/XA + NOUN

       CASE/X NUM/Y GEN/Z ANIM/XA = NOUN TYPE/PHRASE

       CASE/X NUM/Y GEN/Z ANIM/XA

The rule (2), which describes some features of adjective-noun
agreement in Russian, is equivalent to the seventy-two rules
that would be required if each possible case-number-gender-
animateness combinations were referred to explicitly. The
rule can be simply paraphrased as follows: If an adjective
of any case, number, gender, and animateness is immediately
followed by a noun with the same case, number, gender, and
animateness, then the two constituents can be combined to
form a noun phrase having the same case, number, gender, and

animateness. The effect of the rule, when applied to each
of three different pairs of data item constituents, is
displayed in (3).

(3)a. C_1: ADJ CASE/ACC NUM/SING GEN/MASC
ANIM/NO

C_2: NOUN CASE/ACC NUM/SING GEN/MASC
ANIM/NO

C_3: NOUN TYPE/PHRASE CASE/ACC NUM/SING
GEN/MASC ANIM/NO

b. C_1: ADJ CASE/GEN NUM/SING GEN/NEUT
ANIM/$

C_2: NOUN CASE/$ NUM/$ GEN/NEUT ANIM/NO

C_3: NOUN TYPE/PHRASE CASE/GEN NUM/SING
GEN/NEUT ANIM/NO

c. C_1: ADJ CASE/INSTR NUM/SING GEN/FEM
ANIM/$

C_2: NOUN CASE/DAT NUM/SING GEN/FEM
ANIM/NO

C_3: --None defined, due to lack of case
agreement for C_1 and C_2

In (3a), the values of all four attributes specified in (2)
match as required by the repetitions of variables in the
rule. In (3b), there are three instances where a specific
value on one constituent matches a $ or "don't care" value
on the other, yielding by convention a result equivalent to
the specific value. Finally, the constituent pair in (3c)
fails to satisfy the conditions of rule (2) owing to a lack
of agreement in case values; consequently, no higher-order
constituent is produced.

In more formal terms, tags appearing on the left-hand
side of a rule express _tag conditions_, that is, conditions
which the corresponding data items must fulfill if they are
to be permitted to combine into a new data item of the type

described on the right-hand side of the rule. Tags appear-
ing on the right-hand side of a rule specify the tag confi-
guration of the new data item, usually as a function of the
tags of one or both of its components. Tag conditions fall
into various categories corresponding to the different types
of values which an attribute of a rule constituent can assume.
The simplest tag conditions are those involving tags with
constant values; that is, specific values that an attribute
can take on within the object language, such as "accusative"
for "case". A constant is formally defined within the tag
language as any string of six or fewer alphanumeric characters
that begins with one of the alphabetic characters A through W.
If a tag with a constant value appears on the left-hand side
of a rule, as in (4) the corresponding data item must have
the same attribute with the same value (or $) for the rule
to apply. A tag with constant value on the right-hand side
of a rule--for example, the TYPE/PHRASE tag in (2)--simply
indicates that that attribute-value pair is to be assigned
to the new data item which will be produced if the rule
succeeds.

(4)   ADVERB SUBCLS/ADJMOD + ADJ = ADJ

Variable values of an attribute are represented in the
system by alphanumeric strings of six or fewer characters
which begin with one of the alphabetic characters X, Y, or Z.
Unlike constants and $, which appear as values of tags on
both data items and rules, variables may legally serve as
values only on rule tags. If a tag with a variable value
appears on the left-hand side of a rule, the rule applies
only if the corresponding data item has a tag with the same
attribute. If the variable has yet to be defined in a given
attempt to apply the rule, it is defined as the value of the
data item attribute; if it has been previously defined, the
rule tag is interpreted precisely as though it were a tag with
the constant value given in the definition--that is, the

corresponding data item must have the same attribute with
the same value (or $). For example, when rule (2) is applied
to the data item constituents in (3a), at the time the program
processes the CASE/X tag on the ADJ constituent of the rule,
it detects the presence of an undefined variable, scans the
tag string of the ADJ data item until it finds the tag with
the attribute CASE, and defines X as the corresponding value
ACC. After the NUM, GEN, and ANIM tags have been processed
in a similar fashion, resulting in the definition of the
variables Y, Z, and XA as SING, MASC, and NO, respectively,
the program tests for fulfilment of the tag conditions on
the NOUN constituent of the rule. Since the CASE/X tag on
the latter contains the variable X, previously defined as
ACC, the program interprets the tag as equivalent to a CASE/ACC
condition and requires that the NOUN data item have a CASE
tag with the value ACC. When the tag condition testing of
the left-hand side of the rule has been successfully completed,
the program produces a new data item according to the pattern
specified on the right-hand side of the rule, substituting
for each variable the constant value it has been assigned
during processing of the left-hand side. A rule with a
variable on the right-hand side that does not appear on the
left-hand side is not a well-formed statement in the tag
language.

Other tag conditions related to the ones just described
a   those specifying <u>exclusion</u> <u>matches</u>, that is, tag condi-
tions that are fulfilled only if the data item tag has one
or more values which do <u>not</u> occu. on the list of (constant)
values on the corresponding rule tag. Examples of the nota-
tion for the two types of exclusion matches are given in (5).
(5)a.   CASE/ -NOM, ACC
    b.   CASE/ X - NOM
The tag in (5a), whose value is a list of constants preceded
by a minus sign, is interpreted by the program as a condition

requiring that the corresponding data item have a CASE tag
with at least one value that is neither NOM nor ACC. The
program interprets tags with values of the form shown in
(5b) -- a variable, followed by a minus sign, followed by a
list of one or more constants -- in a similar manner: The
data item must have a tag with the same attribute (CASE)
having at least one value distinct from the constants on
the list (here NOM); if this condition is satisfied, the
variable is defined as the list of values on the data item
tag minus any values that also appear on the exclusion list
of the rule tag.

Additional tag operations are illustrated by the rules
in (6) and (7).

(6)   VERB GOV /X + NOUN CASE/X
      = VERB GOVT/1-X ETC/1

(7)   NOUN CASE/NOM NUM/X GEN/Y +VERB MOOD/IND
      PERS/P3 NUM/X GEN/Y GOVT/*
      = MNCLS MOOD/IND ETC/2

As can be seen from examination of the left-hand side of (6),
it is a rule which permits a verb to combine with a noun (or, .
if rule (2) has applied, a noun phrase) on its right, provided
that the GOVT tag of the verb and the CASE tag of the noun
have a value in common (i.e., provided that the noun is in a
case that the verb governs). Unlike the tags on the left-
hand side of the rule, those on the VERB constituent on the
right are of types that have yet to be discussed. The first,
GOVT/1-X, is of the general form ATTR/n-VBL, where ATTR stands
for any attribute name, n is 1 or 2, and VBL stands for any
variable. The program interprets such tags in the following
manner: it copies onto the new data item corresponding to C_3
the ATTR tag from the data item corresponding to C_n (where
n is neither 1 or 2), deleting from the value list of the tag
the value or values corresponding to the variable. Thus, if
(6) were applied to a VERB with the tag GOVT/ACC, DAT followed

by a NOUN with the tag CASE/ACC, the resultant VERB constitute
would have the tag GOVT/DAT. This action would have the
desired effect of preventing the verb from spuriously picking
up more than one accusative object, while still allowing it
to combine with a dative indirect object through a second
application of the same rule.

ETC/1, the second tag on the right-hand side of (6), is
of the form ETC/n, where n is defined as before. Such a tag
is interpreted by the program as an instruction to copy from
the C_n data item onto the C_3 data item all tags whose attri-
butes are not mentioned elsewhere in the rule. Thus, if the
VERB constituent processed by (6) in the preceding illustrat-
tive example had the tags MOOD/IND TENSE/PRES PERS/P3 NUM/SING
GEN/$ (in addition to GOVT/ACC, DAT), the ETC/1 would cause
the resultant C_3 data item to have those five tags in addition
to the GOVT/DAT tag corresponding to GOVT/1-X.

Rule (7) permits a nominative noun (or noun phrase) to
combine with an immediately following indicative verb in the
third person to form a main clause, provided that the two
constituents agree in number and gender and that the verb
has no unsatisfied government potential. The latter condi-
tion is expressed by the GOVT/* tag, which is of the general
form ATTR/*, where, as before, ATTR stands for any attribute
name. Such a tag is interpreted by the program as requiring
that the corresponding data item either have no tag with the
specified attribute or have such a tag with an empty value
list. Thus, rule (7) will accept either intransitive verbs
with no government tag, or transitive verbs whose government
requirements have been fulfilled through successive appli-
cations of (6).

3. Use of the Tag Lanugage in Semantic Analysis

Now that most of the basic features of the current form
of the tag language have been presented, the potentialities
of the language as a tool for carrying out a limited form of
semantic analysis will be briefly considered. In order to
restrict the discussion sufficiently, it will be assumed that
we are concerned not with achieving a full semantic analysis,
but only with checking syntactic analyses (ideally, in the
form of underlying structures) for semantic well-formation.
Further, it will be assumed that this checking is to be
carried out in a manner similar to that of the projection
rule component of Katz and Fodor[1], that is, in the form of
a series of tests and amalgamations proceeding from the
bottom to the top of each tree representing a structural
description of a sentence. The two principal reasons for
this latter choice are the familiarity of the Katz-Fodor
approach and the fact that the tag language is specifically
geared for performing tests and amalgamations in the bottom-
to-top direction.

Katz and Fodor's simplest example is that of the
combination associated with the adjective-noun string "color-
ful ball".. In addition to a syntactic description, in the
form of the sub-tree (8), they assume the presence of corres-
ponding dictionary information (9) for the two lexical items
in the string.

(8)

```
                    Nc
      A ─────────────┴───────────── Nc
      │                             │
   colorful                        ball
```
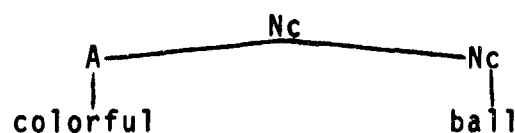
(9) 1. <u>Colorful</u> → Adjective → (Color) → ⟦Abounding in
         contrast or variety of bright color<u>s</u>⟧ ⟨(Physical
         Object) v (Social Activity)⟩
     2. <u>Colorful</u> → Adjective → (Evaluative) → ⟦Having
         distinctive character, vividness, or pictur-
         esquene<u>ss</u>⟧ ⟨(Aesthetic Object) v (Social Activity)⟩
     1. <u>Ball</u> → Noun concrete → (Social Activity) → (Large)
         (Assembly) → ⟦For the purpose of social dancin<u>g</u>⟧
     2. <u>Ball</u> → Noun concrete → (Physical Object) → ⟦Having
         globular shap<u>e</u>⟧
     3. <u>Ball</u> → Noun concrete → (Physical Object) → ⟦Solid
         missile for projection by an engine of wa<u>r</u>⟧

In the dictionary definitions of (9), four distinct types
of information are used in describing lexical strings: 1. <u>syn-
tactic</u> <u>markers</u>, separated form the lexical string by an arrow;
2. a string of <u>semantic</u> <u>markers</u> (each member of which is
surrounded by parentheses), representing that part of the
item's meaning which is systematic for the language; 3. a
<u>distinguisher</u> (in brackets), representing the nonsystematic
part of the item's meaning; and 4., where applicable, a
Boolean function of syntactic amd semantic markers (in angle
brackets) expressing <u>selection</u> <u>restrictions</u> which the item
imposes on other items in certain syntactic combinations. If
distinguishers are omitted, as they presumably can be in a
system aimed only at semantic checking, the dictionary infor-
mation of (9) can be expressed in tag language notation as in
(10). (Here, SEMTYP and SUBCLS represent the principal
semantic and syntactic markers, respectively, and HDSTYP
(head semantic type) reflects selection restrictions that
a particular adjectival modifier imposes on the semantic
type of its noun head.)
(10) a. <u>Colorful</u>
         1. ADJ SEMTYP/COLOR HDSTYP/PHYSOB,
             SOCACT
         2. ADJ SEMTYP/EVAL HDSTYP/AESOBJ,
             SOCACT

B. <u>Ball</u>

1. NOUN SUBCLS/CONC SEMTYP/SOCACT
   SIZE/LARGE SSBTYP/ASSEMB
2. NOUN SUBCLS/CONC SEMTYP/PHYSOB

In their example of semantic analysis, Katz and Fodor
first assign the lexical information (9) to the appropriate
nodes of (8) and then operate on the result using a projection
rule R1, which amalgamates information for a modifier and
its head, provided that the markers of the head satisfy the
selection restrictions specified for the modifier. In order
to operate on the corresponding tag language expressions in
(10) for the prupose of identifying the semantically acceptable
combinations, the tag language rule (11), which is much less
general then R1, can be employed.

(11)   ADJ HDSTYP/X +NOUN SEMTYP/X = NOUN
       SEMTYP/X ETC/2

The combinations from (10) which satisfy (11) are the
following:

(a1, b1) - (dance with bright colors);

(a1, b2) - (physical object with bright colors);

(a2, b1) - (dance with distinctive character).

These are precisely the combinations allowed by the Katz-
Fodor rule, with the exception of (a1, b2), which represents
the merger of two similar combinations resulting from the
elimination of the distinguishers which differentiate Ball 2
and Ball 3 in (9).

4. Extensions of the Language

Although it is at least mildly encouraging to be able
to demonstrate that the current tag language can serve as a
vehicle for a limited form of semantic analysis, a closer
examination of even the very simple example just discussed
points to an area where further extensions of the tag language

would be highly desirable: namely, the representation of
selection restrictions. Because of the nature of the lexi-
cal items in (9), it was possible to make do with a very
simple encoding of the selection restrictions in the data
items of (10) and the rule (11). It should be noted, however,
that although (11) can handle any adjective-noun combination
where only the noun's principal semantic marker is involved
in the selection process, it will fail to apply whenever
additional syntactic or semantic markers are pertinent.
Accordingly, it becomes necessary to write an additional
rule to cover each distinct combination of attributes involved
in selection restrictions.

The nature and magnitude of the problems which arise in
dealing with selection restrictions can be illustrated more
explicitly with reference to the relation of verbs to noun
phrases in their syntactic environments. In (6) we had a
rule which permitted combination of a verb with a noun
provided that the noun was in a case governed by the verb--
in effect, a selection restriction involving only the attri-
bute CASE. If the linguistic facts indicate the desirability
of including an additional restriction for a particular verb--
say, on animateness of its indirect object--it is not possible
simply to add tags for animateness to those for case in (6)
and in the coding for the verb. Instead, new attribute names
must be introduced into the system and new grammar rules must
be written to operate on them. For instance, if it is neces-
sary to indicate for a given verb that it requires an animate
indirect object in the dative case, but that its accusative
direct object is unrestricted with respect to animateness,
this information would have to be recorded for the verb in
the general form indicated in (12), and new subrules (13) and
(14) would have to be introduced into the grammar.

(12)   VERB DOBJCS/ACC DOBJAN/$ IOBJCS/DAT
       IOBJAN/PLUS

(13)   VERB   DOBJCS/X DOBJAN/Y +NOUN CASE/X
       ANIM/Y = VERB ETC/1
(14)   VERB IOBJCS/X IOBJAN/Y +NOUN CASE/X
       ANIM/Y = VERB ETC/1

As soon restrictions on the animateness of subjects, agents, and other verbal complements are described, the proliferation of rules and of distinct names for the same attribute (e.g., CASE, IOBJCS, DOBJCS) will increase. Simi-lar effects can be anticipated in dealing with selection restrictions in other segments of the grammar. The result will not only be esthetically unpleasing from a linguistic point of view, but will also have serious practical conseq-quences in terms of a very substantial increase in space and time requirements for processing the grammar, whether manually or automatically.

A potential solution to this problem which currently appears attractive involves extension of the tag language through the introduction of "super attributes" that have strings of tags as their values. The present tag notation permits reference to all possible combinations of individual values of a specific set of attributes by means of a single rule where the tag for each of the pertinent attributes has a variable as its value. The proposed extended notation employs a similar device: all tag strings that have a parti-cular super attribute are referred to in a single rule by the appropriate "super tag" with a variable value. Thus, in place of (12), (13), and (14), we introduce the data item coding (15) and the rule (16) with the super attribute SELRES, where the super tags are distinguished from regular tags by the double slashes flanking their value fields.

(15) VERB SELRES//CASE/ACC, CASE/DAT ANIM/PLUS//
(16) VERB SELRES//X// +NOUN //X//
                    = VERB SELRES//1-X//ETC/1

Since much additional work remains to be done in exploring the implications of introducing super tags into the tag language system, the notational conventions employed in (15) and (16) are extremely tentative in nature. By the same token, it is clear that programming of routines for interpreting the new notation lies still farther in the future. Nevertheless, on the basis of present evidence, it seems equally clear that such an extension of the current tag language will be necessary to provide a capacity both to perform more effective syntactic analysis and to carry out extensive semantic checking.

REFERENCES

1. Katz, J. J. and Fodor, J. A., "The Structure of a Semantic Theory", Language, Vol. 39, No. 2 (1963), pp. 170-210.

2. An Introduction to COMIT Programming, The Research Laboratory of Electronics and the Computation Center, MIT (1961).

3. Plath, W. J., "Multiple-Path Syntactic Analysis of Russian", Mathematical Linguistics and Automatic Translation, Rpt. No. NSF-12, Harvard Computation Laboratory (1963).

4. Robinson, J. J., Endocentric Constructions and the Cocke Parsing Logic, RAND Paper P-3101 (1964).

DISCUSSION

YNGVE: I have a few comments. I think they all come under the general question of. Could you do all this in COMIT?

PLATH: I'd say "probably yes."

YNGVE: From there on it's a question of ifs, ands and buts, and I thought I would like to discuss some of the ifs, ands and buts.

There are several things that enter into a decision as to what language to use and how to program something. One of them is the ease of programming; that is, the convenience, the aesthetic appeal and so on that is involved, and this is a very important aspect for the person that is actually dealing at the top level with a program. Then there is also, of course, the question of the overall programming time, including all of the other people that work on it. Then, in addition, there are questions of storage space and running time of the programs.

If you are to do these sorts of things in COMIT, and I would be inclined to do that myself, first of all you could program it directly -- that is the operations; the ideas behind what you are trying to do you could program directly in COMIT.

PLATH: I am quite aware of that.

YNGVE: If you preferred to write your programs in a slightly different notation because of convenience and aesthetic appeal there are two general methods open to you if you decide to base your work on COMIT. One is to write a compiler in COMIT that translates from your new notation into COMIT, and then that is run.

In COMIT, too, the system facilities are very conveniently available for running a series of jobs like that, so that you don't essentially see the two-steppedness of the process, which was not true in COMIT 1.

The other thing that you can do, and you can do this concurrently, is to make use of your COMIT features, allowing COMIT to call machine language routines which would give you, for example, different subscript operations, different kinds of merging. This is a new facility in COMIT 2 and COMIT 2 has not been advertised or distributed. However, I can tell the group here that if anyone wants to use it we can send it prior to SHARE distribution. It is now in a state where it is practically debugged, and we are immediately willing privately to sent it to any people who seriously want to use it for linguistic data processing. It runs on the 7040, 7044, 709, 7090, and 7094.

PLATH: Since COMIT 2 wasn't available at the time we were working on this, we weren't able to consider its features.

YNGVE: Everything I say, except this machine language facility, you can do with COMIT 1, like creating the compiler and so on.

PLATH: This thing is somewhat different in the sense that rather than imbedding machine language routines in something like COMIT, we in effect have tag language grammars imbedded in a machine language program. We turned things upside down.

YNGVE: Yes. That way you may achieve advantages in speed of running, and also retain a considerable amount of the convenience.

ROSS: I can see how conceivably, barring all the ails, the failings, of a simple feature or componential analysis which was played out by Bar-Hillel, I can conceive of some use of

this for semantics. I can not conceive of any use at all for
a symbol like the output of 2a(3), noun-type phrase, and all
these things. You don't have to mark noun phrases as to
whether they are genitive or singular or anything like that,
or at least I know of no case where you even need this informa-
tion.

The kind of selectional restrictions and so forth that
you need can be stated in terms of the features that they
have now.

PLATH: In terms of this algorithm, it depends on how you
are parsing. If you simply have a string of adjectives and
a noun and you combine more or less in this order, one of the
things you want to know in order to decide whether or not to
perform a combination, or whether a combination is legitimate
or not, is what was the case of this noun. But somehow, since
in the mechanism of the program once you are combining another
adjective, or trying to combine it with this combination, you
are dealing not with this directly, but with this combination
of it. You have to somehow pass on the crucial information up
to this node. It is just part of the sequence of operations
here. I don't think it has any deep linguitic significance
in any sense. It isn't meant to.

GARVIN: Wouldn't you need it for linguistic purposes if you
have a single plural adjective with two singular nouns? The
resultant phrase is presumably a plural nominal phrase, a fact
which is not shown by the grammar code of either of the consti-
tuent nouns.

PLATH: Yes.

XIII.

AN APPROACH TO THE SEMANTICS OF PREPOSITIONS*

Ernst von Glasersfeld

Instituto di Documentazione,
dell'Associazione Meccanica Italiana
Milan, Italy

As a preface to my paper I should like to remark on
something that became noticeable during the sessions of
this meeting. The word "intuitive" has cropped up quite
a number of times, and nearly every time a speaker used
it he did so almost with a sense of guilt.   I don't un-
derstand why this should be necessary. Language, to my
mind, is an extremely intuitive arrangement of things,
intuitive in its production and intuitive in its inter-
pretation. This is not to say that language does not in-
clude logical functions and logical implications, but it
embraces very much more. For instance interpretations
that are "correct" merely because they are much more pro-
bable than others, given our experience of the world we
live in.

When a human being uses language he never actually
calculates these probabilities - he assesses them impres-
sionistically or, if you like, he makes guided guesses.

In this connection there is a suggestion I should like
to make and I assure you that I don't mean to be nasty
in any way: Would it not be a good thing if the master
logicians, who never miss a formal slip or an illogical-
ity in the empirical linguist's attempts to unravel language,

were to apply their minds to the very real illogicalities
of natural language? - If they did, I am confident, they would
soon come up with finds that could be a help to all of us.

Among tranditional linguists and grammarians the title
of this paper may cause some bewilderment. Prepositions,
for a long time, have been thought of as 'function words'
and considered to have no meaning in the sense in which nouns,
adjectives, etc., have meaning. That this view is not
altogether a thing of the past is shown by the recurrence of
the statement that prepositions are not 'important' words.
This view probably was and is most firmly supported by docu-
mentalists who are approaching the problems of information
retrieval by means of 'key-words', 'content-words', 'micro-
glossaries', etc.; even in that field, however, a number of
research groups have come to the conclusion that the relations
obtaining between the words of a given text are often essential
parts of the content expressed by it and, consequently, these
groups have tried, in one way or another, to make their system
sensitive to relations (1). This has led them to consider
more closely, among other things, the various types of relation
that can be expressed by prepositons.

Linguistic research that in some way aims at a workable
procedure for machine translation comes up against problems
created by prepositions the moment it examines a natural text,
i.e. a text that was not written for a translation experiment*.
We are all familiar with output from Russian-English translation
programs where, for Russian prepositions met in the original
text, the print-out displays more or less numerous selections of
English 'alternatives' among which the reader is supposed
to choose; and it is perhaps not always pointed out with

---

*We are certainly not the only group who has become aware
of this; the first, in my knowledge, was Silvio Ceccato's (2);
since then also members of Sydney Lamb's school have approached
the problem (3).

sufficient urgency that such a choice among 'alternative' prepositions is neither a question of mere style nor, frequently, is it a choice made obvious by the context, especially if all one has to go on is the _translated_ text.

Since prepositions, as their primary function, express relations between other elements of a sentence, some might prefer to consider a study of these relations as belonging to syntax rather than to semantics. For the correlational grammar we are working on, this makes no difference whatsoever, because conventional syntax and semantics are to a considerable extent amalgamated in it. However, having heard Mr. Pankowicz's splendid empirical definition of semantics ("The study of what is supposed to remain unchanged when we translate an expression, phrase, or text from one language into another") I am confirmed in considering preposition analysis as belonging to the field of semantics; because the analysis of prepositions, and of relations in general, has the very purpose of making sure that the relations expressed in a given sentence remain unchanged when the sentence is translated into another language.

Without going into any philosophical discussion about terms such as "meaning", "synonymy" and "ambiguity", I think it should be clear that the sentence

(A) - There are many books about John's house -

is ambiguous (in a sense that is of paramount importance to language analysis and machine translation) and, further, that the ambiguity in this case springs exclusively from the fact that the preposition "about" has the nasty capacity to express more than one relation.

If we circumscribe the two relations that may be operative in this sentence, we can distinguish:

a) the relation obtaining between a 'semantic' object*
   (such as it is designated by "book", "story", "play",
   etc.) and its subject matter, and
b) the spatial relation obtaining between several spatially
   limited objects and an enclosed space within which they
   are located.

If such a sentence occurs in a document that has to
be summarized or in any way analyzed for documentation purposes,
it may become necessary to resolve its ambiguity. If it occurs
in a text that has to be translated, it is indispensable that
the ambiguity be resolved, because different relations, as a
rule, require different output.**

Before embarking on a discussion of how one might handle
the specific relations expressed by prepositions I should
like to stress that it is by no means only prepositions that
create relational problems, but also a number of syntactic
constructions that have nothing to do with prepositions at
all. The sentence:

(B) - The man hit the ball -

                                    has cropped up as an example in
quite a number of books and research reports, mostly, I suppose,
because it seems to be fairly straightforward; that is to say
a normal English-speaker would not consider it ambiguous at
first sight. If it has to be translated into German, however,
we may become aware of the fact that we cannot really be sure
that this sentence means unless we get some additional informa-
tion about the situation to which it refers.

---

*Elinor Charney, in her paper, used a much better name
for this class of object: 'communication-bearing objects' -
we shall gladly borrow it from her in the future.

**In the case of sentence (A) being translated into German
we should get on the one hand "es gibt viele Bücher über Johns
Haus", if the author of the sentence meant to say that the books
had been written about John's house; and on the other, "viele
Bücher liegen in Johns Haus umher" if the books were supposed
to be lying about in John's house.

If there was some mention of golf, tennis, cricket,
or baseball, or if the man was last heard of with a club,
a racquet, a bat, or even just a stick in his hand, we
should have no qualms about putting down the translation
"der Mann schlug den Ball"; if, however, we last saw our
man with a gun in his hand and at the counter of a shoot-
ing booth, we should translate "der Mann traf den Ball";
and finally, - unlikely, but surely possible -, if we
had watched a juggler perched on a ladder miss a catch,
lose his balance and come crashing down, we should be
inclined to translate "der Mann schlug auf den Ball auf".
All this can, of course, be put down to an inherent ambi-
guity of the English verb "to hit" - but, and this is what
I should like to stress in this context, it is essential-
ly a problem of relations, i.e. of different relations
obtaining between the man and the ball.

One of the major differences between conventional
sentence analysis and the kind we are trying to perfect
in our project is that we set out to map (i.e. to isolate,
classify and code) the relations which can be conveyed
by language, and we try to do this regardless of whether
the relations are among those that are usually described as
syntactic or not.

At this point, as a rule, two objections are raised
against our approach. The first boils down to the ac-
cusation that by considering all sorts of relation in
our "correlation grammar" we aid and abet the general
confusion of the terms "syntax" and "semantics". Seeing
how much some of the most venerable philosophers have
done to foster that confusion, we are unable to feel
very guilty about this. The second objection, however,
is very serious. It is often couched in different terms,
but essentially it amounts to this: the relations that can
be expressed in natural language are so diverse and so

many that any attempt to map them all is bound to fail -
and even if it succeeded, no computer would be large enough
to handle them (4). This worries us a great deal because,
although we do not agree with the dismal conclusions, we
know only too well how correct the premises are. The
number of relations to be isolated, classified and coded
is, indeed, enormous and we are painfully aware of the fact
that what we have done until now is only a very small fraction
of what has to be done. As to the capacity of computers, we
see no reason to be pessimistic; advance in computer design
has been and presumably will be so much faster than ours
that it seems a rather safe assumption that by the time we
have mapped linguistic relations machines will be able to
handle more than we have to put in. It may still be important
to produce a really suitable and economical program - but on
that count, too, we feel no apprehension. As to the enormous
amount of analytical work that remains to be done, there is
at least one consoling feature: it does not have to be done
in one fell swoop. We are concerned with language analysis
for the specific purpose of machine translation, not with
mapping the semantic universe of the human mind.

Let me try to explain the difference I want to make.
The relations the human mind posits between items it strings
together in its thinking are indeed astronomically many,
and although I do not believe that their number is infinite
I have no doubt that it would take a very large research
team something like a lifetime to catalogue them all. On
the other hand, the languages human beings use to communicate
their thoughts have a certain amount in common. In particular
the languages with which we are at present concerned* have
a great deal in common. And what they have in common (in their

_____

*English, Italian, French and German

ways and means of conveying relations) does not have to be
broken up any further for the purpose of translation.

In practice this means that certain English expressions,
although ambiguous as to the relations they convey, do not
require those relations to be treated individually, because
the languages into which we want to translate happen to offer
expressions which are correspondingly ambiguous.

For instance, if the sentence
(C) - I'll do it in twenty minutes -
is to be translated into German, we can disregard the fact
that the preposition "in", in this context, conveys two
different relations - i.e. (a) the activity will be terminated
within twenty minutes, and (b) the activity will take place
after twenty minutes -, because the German "in", in a similar
context, is ambiguous in precisely the same way. When we
come to translating the sentence into Italian, we could make
a fairly strong case for still disregarding the ambiguity,
because the Italian "in", at least colloquially, is used in
the same way; a purist, however, would object that the relation
(b) should, in Italian, be expressed by the preposition "fra";
and therefore, to have the output really clean, we would have
to split the two relations in our input analysis.

The example is somewhat trivial, but it may help to show
what we mean when we say that the depth of our analysis of an
English preposition ('explicit correlator') is determined by
the output requirements of the languages with which we are
dealing (5).

In practice we try to split and isolate the relations
conveyed by a preposition whenever we find that one (or
more) or our output languages requires a distinction. Some
of these distinctions are undoubtedly of the kind which Chomsky (6)
can account for by his system of transformations. For instance,
our empirical finding that the English "by" in the sentence

(D) - George was betrayed by his stammer -

requires the German preposition "durch",while in the sentence

(E) - George was recognized by his stammer -

it requires the German preposition "an", can be neatly discri-
minated and substantiated by the demonstration that sentence (D)
is a transform of

(D') - his stammer betrayed George -

whereas no similar transformation is possible for sentence (E).

There are, however, other distinctions which are not
immediately explicable by means of transformation. They are,
I think, of the kind which Charles Fillmore is planning to
handle by means of his Semantic Entailment Rules for an
integrated generative grammar (7).

An example of this kind of distinction, which we have
found particularly tiresome, crops up with certain 'causal'
or 'instrumental' uses of the English "by". As far as the
preposition is concerned, to an English-speaker, a Frenchman,
or an Italian, it makes no difference whether a man be killed
by a stroke of lightening, an arrow, or a shot. For a German,
however, there are some intricate considerations to be made
before he can decide on the proper preposition. (I am far from
certain that my present analysis is correct or even applicable
within our very limited vocabulary; so I present it here as an
illustration of our method rather than as a final result) -
The distinctions to be made concern the intentionality of the
act and, on a further level, whether the result of the act can
be considered its direct consequence or merely a corollary.
Thus, in the German-speaking world, in spite of autochthonous
gods notorious for their thunderbolts, death from lighting is
considered a direct consequence, but not the realisation of

someone's intention; the preposition, therefore, is "von". In
the case of the arrow, death, again, is the direct result, but
the archer may or may not have aimed at the man; if there was
intention, the preposition to choose is "durch", if not, it
is "von". A shot, finally, seems to be considered intentional
under all circumstances and it is linked to its result by the
preposition "durch". It seems coherent with this pattern of
ideas that, for a German-speaker, it is always unintentional
when someone is grazed by a shot; and in this case the preposition
to choose is "von".

The tiresome complication arises when we are concerned with
acts which are not intentional and with results which are not
considered a direct consequence. For instance, the sentence

(F) - Smith was ruined by the economic crisis -

requires the German "durch". In the preceding examples this
preposition seemed to convey intentionality; here, however,
it is scarcely plausible to maintain that the economic crisis
intentionally ruined Smith. So we fall back and say, not only
is there no intention in this case, but the result is not
considered a direct consequence either. This allows us to set
up the schema:

intentional	direct consequence	= "durch"
intentional	corollary	= "von"
unintentional	direct consequence	= "von"
unintentional	corollary	= "durch"

Although this looks very neat and satifactory, it by no
means solves the whole problem. Above all there remains the
difficulty of deciding in certain cases what is to be considered
a 'direct consequence' of an act or event, and what not. In
many instances it would seem that what we tentatively called
'direct' consequence, is something that is generally considered
a normal consequence of the particular thing, act, or event

mentioned. Thus it would be <u>normal</u> for a wind to blow things
away, but not <u>normal</u> for it to open windows; and this, (the
wind not being endowed with intention) is corroborated by
the German use of "von" in the first and of "durch" in the
second case.

Obviously these are subtle and at times hazy differentiations
and we do not for a moment declude ourselves that they could be
called scientific. They are, however, useful insofar as they
help us to isolate and to bring into some kind of system the
often very elusive factors that determine differences in the
output springing from one and the same English preposition.

Besides, we have come to realise during this research that
the distinctions we have to make - especially insofar as they force
us to discriminate words according to their capability or inability
to function as terms of a given relation - will supply something
like a skeleton for a general semantic classification. This
is still an impression rather than a definite conclusion, because
we have as yet not ordered the data in a systematic way; our
impression, however, fully corroborates what James H. White
states at the end of his article on 'sememic analysis', i.e. that
analysis of prepositions constitutes "an excellent jumping off
point for a sememic analysis of the rest of the language" (3).

I have dwelt at great length on this one problem of isolating
specific relations because it does show the difficulties that
have to be overcome and, I hope, also the kind of reasoning we
try to apply. By comparison, the problem of classifying and
coding the relations is very simple. It can be summarised very
briefly.

We have so far made a preliminary analysis of twenty English
prepositions (about, after, at, between, but, by, down, for, from,
in, like, of, on, since, than, through, to, under, up, with) and
a thorough analysis of two of them, "about" and "by". The analysis
of "about" has given rise to 32 relations, that is, we have had

to split the relations conveyed by this preposition into 32
individually characterised ones in order to assure that when-
ever one of these is recognised in an English sentence we
can directly indicate the output that corresponds to it in
Italian, French and German*.

The analysis of "by" has given rise to 34 relations. On
the basis of the preliminary analysis we expect the most
ambiguous English prepositions - "of", "to", "in", and "on" -
to yield a maximum of about 80 or 90 relations, and if the
worst comes to the worst, not much over one hundred.

In our first, extremely crude coding system, each preposition
is identified by a code number of three places.

The limited vocabulary with which we are working consists
of approximately 500 inflected words and each one of these is
examined for its posibilities of entering as a first or as a
second term (first or second 'correlatum') of the coded relations.
Taking as an example the relations expressed by the preposition
"about" in sentence (A), we proceed as follows:

---

*This does not mean that for every occurrence of "about"
we determine one and only one output in the other languages;
it merely means that we account for the ambiguities and have
an output ready for each possible meaning. Ambiguities of the
kind illustrated by the example (A) - and there are a great
many sentences of this kind in natural texts - can be resolved
neither by a human translator nor, consequently, by a machine
program, unless additional information from a wider context
outside the one sentence is made available (Without context they
can, at best, be approached by a probability rating). This
constitutes, indeed, a serious problem not only for the
resolution of relational ambiguities but also for the resolution
of many ordinary lexical ambiguities. In our view, it can be
approached only after an analysis procedure for single sentences
has been successfully implemented; and by successfully implemented
we mean that the procedure produces all analyses (or interpretations)
that are possible for the single sentence. For only if we have
all these interpretations to hand, can we proceed to eliminate
some of them on the basis of information supplied by the wider
context.

1) we take each item in our vocabulary and ask whether it
   can possibly occur as the first term of the particular
   relation we are considering at the moment (in our case
   003/031, the relation obtaining between several spatially
   limited objects and an enclosed space within which they
   are located; cf. footnote on p. 2). If we find that it
   can, the item is assigned the code number of that relation
   on its 'word-card', and also the indication that it can
   function as first term of that relation.

2) Each item is then examined for its possibility as the
   second term of that relation; wherever the possibility
   exists, it is again recorded on the item's 'word-card'.

3) The same examination is repeated for the next relation
   (e.g. 003/191, the relation obtaining between a 'semantic'
   object and its subject matter; cf. footnote on p. 2).

   Having completed this, we find that the word-cards cor-
responding to the 500 items in our vocabulary show the following
distribution of relation indices. Index 003/031 occurs with
function 1 (i.e. possible 1st term of the relation) on the
items:

books	nines
cakes	ones
cans	pieces
glasses	tables
hands	threes
houses	towns
lemons	twos
letters	

Index 003/031 occurs with function 2 (i.e. possible 2nd term
of the relation on the items:

house	houses
mine	mines
town	towns

Index 003/191 occurs with function 1 on the items:

answer	answers
book	books
letter	letters
question	questions
reading	readings
saying	sayings
story	stories

Index 003/191 occurs with function 2 on all items which
can function as accusative object of a verb (in our vocabulary
they amount to a total of 212 items including the 30 that occur
in the above three lists).

Given this index distribution, our analysis procedure*
recognizes, for instance that the "about" in the sentence

(G) - there are cans about the house -

expresses relation 003/031, because "cans" and "house" show
the index of that relation with the functions corresponding
to their position in the text; and relation 003/191 is not
found in this sentence, because "cans" does not bear that
index with function one (as would be required given the position
of the word relative to "about").

If the analysis procedure is applied to sentence (A)

(A) - There are many books about John's house -

it recognizes the ambiguity of "about" in this case, since the
item "books" bears both the indices 003/031 and 003/191
with function 1, and the item "house" bears both these indices
with function 2.

Finally, in the sentence

(H) - There is a story about John's house -

the analysis procedure recognizes that "about" expresses

---

*A full description of this analysis procedure is contained
in the two reports listed under Nos. 5 and 8 in the bibliography.

relation 003/191, because "story" and "house" bear that index
with the required functions; and it will not find relation 003/031,
because "story" does not bear index 003/031 with function 1
(as would be required, given its position relative to "about").

This crude and somewhat naive example should also make
clear that our relational analysis of prepositions does not
resolve real ambiguities. We merely claim that it takes account
of them, brings them up to the surface, as it were, and, by
differentiating and coding the various possible interpretations,
prepares the ground for their eventual elimination by means of
such other information as can be gathered from the wider context;
and we also claim that the system helps to avoid a considerable
number of what I should call pseudo-ambiguities. This is a
delicate point. Professional linguists could be maliciously
described as people who are always able to find counter-examples
to the rules their colleagues make. Such examples can always be
found and often they can be made to sound quite plausible. If
a short story writer visited John's house and discovers later
that he left some of his manuscipts there, he might conceivably
telephone John and ask: "Did you by any chance find my stories
about your house?" - And in that case our analysis procedure
would miss his meaning because in our system the word "stories"
is not indexed as designating the kind of object that you can
leave about a house*.

Failures of this kind do not discourage me.
The writer who formulated his question in that manner is simply
asking to be misunderstood. He is using his words misleadingly,
because he strings them together in a way that must give rise
to a common and obvious interpretation - while the interpretation
he actually wants to cause in the receiver is another one. Human

---

*Note that the second meaning of "stories" (=floors) does
not get that index either, because the index is assigned only
to words designating objects whose location is not predetermined
or implicit.

receivers are, as a rule, extremely lenient and flexible with
regard to that sort of linguistic irresponsibility (in case of
doubt they trust their knowledge about an experiential situation
more than the linguistic formulation that refers to it); they
try to understand as best they can, and their best is pretty
good. Nevertheless there are, I believe, certain limits of
improbability beyond which a user of language should not place
the interpretation of his message which he wants the receiver
to make - unless the writer be a poet who more or less deliberately
uses puns or hermetic formulations as a literary instrument.

For all of us who are trying to train computers in the
use of language, these limits of probability or improbability
are indispensable. For the more we water down linguistic rules
and relax restrictions in order to allow for improbable relations
and constructions, the less univocal the interpretation will
be in the many cases where the probability of one interpretation
is so great that, for the human receiver, it amounts to certainty.
The question, therefore, is not whether to set up restrictions
or not, but it is <u>where</u> to set them up.

In the course of our research on relations we have found
that there are different kinds of semantic improbability, oddity,
or impossibility. By and large we agree with the distinctions
made by other investigators (9). Several types of semantic oddity
problems seem amenable only to a sophisticated system of probabi-
lity ratings. A sentence such as "there were several hands about
the house" may look odd at first sight; but apart from the fact
that, say, "farm hands" may have been mentioned just before,
the sentence (with the ordinary meaning of "hands") could
conceivably occur in a horror story. So, at best, we can say
that it is somewhat improbable. "There were several towns about
the house" would seem more than a little odd, but would we be
justified in excluding it altogether? - I think not. It would
only need some introduction of the kind: "Last night I dreamt
that No. 10 Downing Street had grown to an enormous size". And

since accounts of dreams are not a negligible quantity in the literature of psychology we cannot even exclude oddities of that kind by saying that we are interested in scienfific texts only. So, again, we can merely say that the sentence is improbable, more so than the previous one, but not impossible.

There is, however, one thing in all these sentences that we can rule out absolutely: the pseudo-ambiguity that arises as long as prepositions are taken indiscriminately as function words. The "about" in the above examples (and in sentence G) cannot express the relation 003/191, no matter how we introduce, preface, or transform these sentences. Given the system of relation-indices, our correlational analysis procedure eliminates this type of pseudo-ambiguity, because the impossible relation does not even come up as a tentative interpretation, and this elimination can be applied to a great many prepositional constructions and to all the prepositions we have examined.

REFERENCES

1. Jessical Melton, "A Use for the Techniques of Structural
    Linguistics in Documentation Research," Technical
    Report No. 4, Center for Documentation and Communi-
    cation Research, Western Reserve University, Cleve-
    land, Ohio.

    Jean Claude Gardin, "SYNTOL," Rutgers Series on Systems
    for the Intellectual Organization of Information,
    Vol. II, The Rutgers University Press, New Brunswick,
    New Jersey, 1965.

    Maurice Coyaud, "Document Automatic Indexing with the Help
    of Semantic Information," Section d'Automatique
    Documentaire, C.N.R.S., Paris, 1965.

2. Silvio Ceccato et al.; "Linguistic Analysis and Program-
    ming for Mechanical Translation," Technical Report
    RADC TR-60-18, Feltrinelli Editore, Milan, 1960.

3. James H. White, "The Methodology of Sememic Analysis with
    Special Application to the English Preposition,"
    Mechanical Translation, Vol. 8, No. 1, August 1964.

4. Michael A. Arbib, "Notes on a Partial Survey of Cyber-
    netics in Europe and the USSR," Final Report,
    Contract AF 49 (638)-1446, Directorate of Information
    Sciences, AFOSR, Washington, May 1965.

5. Ernst v. Glasersfeld, "A Project for Automatic Sentences
    Analysis," Beiträge zur Sprachkunde und Informations-
    verarbeitung, No. 4, Munich, 1964.

    E. v. Glasersfeld, P.P. Pisani, J. Burns, B. Notarmarco,
    "Automatic English Sentence Analysis," Final Report,
    Grant AF EOAR 64-54, IDAMI Language Research Section,
    Milan, June, 1965.

    Jehane Burns, "English Prepositions in Automatic Trans-
    lation," Beiträge zur Sprachkunde und Informations-
    verarbeitung, No. 7, Munich, 1965.

6. Noam Chomsky, Aspects of the Theory of Syntax, The M.I.T.
    Press, Cambridge, Massachusetts, 1965.

    J.J. Katz, P.M. Postal, "An Integrated Theory of Linguistic
    Descriptions," Research Monograph No. 26, The M.I.T.
    Press, Cambridge, Massachusetts, 1964.

7. Charles J. Fillmore, "Entailment Rules in a Semantic Theory," in _Project on Linguistic Analysis_, Report No. 10, pp. 60-82, The Ohio State University Research Foundation, May 1965.

8. E.v. Giasersfeld, P.P. Pisani, J. Burns, "Multistore, a Procedure for Correlational Analysis," _Report_ T-10, IDAMI Language Research Section, Milan, January 1965.

9. Angus McIntosh, "Patterns and Ranges," _Language_, Vol. 37 No. 3, July-September 1961.

   Peter Kugel, "Some Remarks on the Structure of Semantic Theories," paper presented at the _2nd_ _AMTCL_ _Meeting_, Bloomington, Indiana, July 1964.

   Noam Chomsky, _op_. _cit_., p. 95 and p. 152.

DISCUSSION

ROSS: First of all, a couple of purely syntactic comments.
I am very interested in the paper. The phrase "There are
several books about the house" is, of course, disambiguated
syntactically. In one case "books about the house" is a
noun phrase and in the other case it is not. "There are
several books" and "about the house" really should come from
something like "Several books are about the house" as part
of the predicate.

VON GLASERSFELD: How do you spot this syntactical difference?
That is precisely our problem.

ROSS: Well, I think the way you are going about it is
precisely right.

VON GLASERSFELD: That is why I said I don't speak about the
term "semantic." If you want to call these relations syntactic
I perfectly happy. They obviously embrace what you mean when
you talk about noun phrases and predicates and so on. They must
embrace it if they want to understand natural language, because
traditional syntactic terms, after all, have been useful for,
I don't know, four or five thousand years in teaching languages.
My contention is merely that they are not complete; not complete
in the sense that if you want to explain language to a computer
that hasn't got the upbringing and the experimental knowledge
of a child, you have to explain far more.

ROSS: The second point is about the case with "in" -- "in five
minutes." As I understood you, you seemed to couple this with
the point that "The man is in the house" as opposed to "The man
is in the room" -- whether or not to treat that as the same.

VON GLASERSFELD: No. That is an entirely different "in."
It is accidental that one example came after the other.

ULLMANN: Three small points. In the first place I can bear
out what you said on "en" and "dans." It is perfectly true.
"En" means that he will do it within a space of five minutes,
and "dans" that he will start in five minutes.

On the fundamental point, to which you seem to come back
and which seems to be worrying you, whether this is semantics
or syntax, I think the root of this is that here we are dealing
perhaps with a new type of word. Prepositions seem to be a
new type of word which are nowadays called "form words" whose
function is grammatical rather than lexical, and that is why
you were wondering whether this is syntax or s mantics.

But I don't see any opposition between these two. To my
mind, both lexes and syntax have a semantic component and a
form of morphological component, whatever that means. What you
are doing very well is syntactic semantics.

One final point; a question, rather, which really follows
what I was saying. This is really a matter of terminology.
I am just wondering, and this would really be a very important
matter, whether these form words, like prepositions exhibit the
same sort of features? Have you been able to find some sort of
underlying unity, some sort of common substratum behind these
twenty or thirty different uses of the same preposition or not?

VON GLASERSFELD: In different languages, or in one and the same?

ULLMANN: In any particular language, at the moment. Are these
homonyms, so to speak?

VON GLASERSFELD: We discovered, or I think it is known by every
user of the language, that a preposition like "in" has certain
spheres to which it applies. One you can call spatial and the

other you can call temporal, a third one you can call modal.
It seems obvious to me that it must be like that. It is
like that in all of the languages that we deal with. But I
might add to this that we at the moment -- and this is until
we shall have finished the comlete analysis of at least these
twenty prepositions -- don't try to categorize the descriptions
in any way. We make them as they come, as we find them useful
to discriminate the word itemson the one hand and the prepo-
sitional relations on the other. When we have finished we shall
try to see what kind of an order we can bring on the one hand
into the description of the word groups; on the other, into
the description of the relations.

MACDONALD: This is a topic that I have been very much interested
in and I could probably go on for several hours. I would like
to make one or two small points.

It seems to me that any linguistic form has two values
that might be called an interlinguistic value and an extra-
linguistic value. That is, it may have two. Many of them have
only an interlinguistic value. This is true of certain prepo-
sitions. These are the prepositions whose usage is determined
not by their object or by the qeneral syntactic structure of
the sentence, but by some particular item in the sentence.

For example, "consist" requires "of" and "depends" requires
"on." If you set those aside, then I think you will find that
prepositional structures, at least in English, function generally
as adverbs. An analysis of your adverbs will produce a certain
number of sub-classes. I have at least seven.

The prepositional phrases function in one or the other
of these sub-classes and, in fact, are in direct contrast with
certain adverbs where there is no preposition involved. That
is, "on the table" operates in much the same way as "here" or
"there," and "here" or "there" should really, perhaps, be
described as being a type of prepositional structure.

In these cases I can't find that you can determine which
function the preposition is fulfilling by categorizing the
object in much the same way as you suggest, and once you have
determined that it is fulfilling the work of, say, a Class 1
adverb, then there is very little difficulty in determining
the semantic value.

Now then, in the case of "There are many books about the
house," the sentence is ambiguous under any circumstances.
I mean, if you give just the sentence without anything before
it or after it. And therefore you can not really expect any
sort of semantic organization to resolve that ambiguity with
just that much context.

In fact, I think there are three possible things to be
considered here. There could be books which are written about
the house; there would be books which could be about the house
and inside the house; and, much less probably, there are books
which could be about the house and outside the house.

VON GLASERSFELD: We have the latter relation too.

MACDONALD: I think it would be preferable if they were
closely linked with the adverbs, or the adverb classes,
because they perform much the same function, as it seems to
me.

VON GLASERSFELD: I have no particular opinion about whether
it is better to treat prepositions and prepositional phrases
as kinds of adverbs or not. The reason why we do not treat
them as adverbs, at least not in prepositional constructions of
this kind, is that they fit much better into the analysis
procedure that we have designed. We would have to alter our
procedure radically to fit in prepositional adverbial phrases
of that kind. An alteration can be made, but at the moment I
don't feel like making it because I want to see how far I can

with this way of combining words by means of prepositions and
using prepositions in much the same way as the other syntactic
functions.

    This idea springs from Ceccato's school. I don't in any
way believe that it is the only possible one. It seems to
me to have a number of practical advantages in so far as the
machine procedure of analysis is concerned. That is all.

ROSS: I think the point Mr. MacDonald made about "consists"
is a good one, and generalizes. I don't think it is a fruitful
idea to try to analyze the meaning of "of" in a phrase like
"consists of". Furthermore, there are other prepositions which
are syntactically determined, such as "I ave lived here since
Christmas, but "I have lived here for two weeks." I think it
would be foolish to try to consider "since" and "for" as
being two different prepositions. The same applies to things
like "He came at four o'clock; he come on Tuesday."

VON GLASERSFELD: I don't pretend that it's right or wrong or
anything like that. If you can show me that by lumping dif-
ferent things like that you can translate I shall be very
happy. What was your example? "Coming on Tuesday", "Coming
at four o'clock", "Coming in the summer of next year" --
try to translate these things into the three languages I have
mentioned. You will find that the output does not respect your
idea of unity there. That is why I break it up.

MACDONALD: I think the difficulty is that you are working from
a preposition to a preposition. If you consider that those three
prepositions all belong to a class of time adverbs and that
they express points of time, and then build up a different
system for whatever language you are going into as to how that
language expresses point of time, you would get from one to
the other without any difficulty, and it wouldn't be a matter

of saying "in" is to be translated in this way, but of
saying "in" in this case expresses point of time for English;
point of time in Italian is done this way.

VON GLASERSFELD: I think this is a beautiful illusion, that
something like point of time can be generalized to that extent.
Select your expressions in English that express what you call
point of time and translate them into the three languages.

    Now, as I say, I don't say this the only way of doing
it, but I try to finish it to see how far it will get us. I
have no claim of perfection or anything like that. But one
thing that has come out in this meeting, I feel very strongly,
is that everybody criticizes straightaway a practical attempt,
as though it were a mistake to finish one approach without
the glorious idea that it is the only one and that it is the
right one. But let us do some field work, even if the theory
underlying it in the end will prove wrong. The kinds of splits
that I make will be extremely useful to you for sorting out
your adverbial expressions.

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Wayne State University Detroit, Michigan 48202	Unclassified
	2b. GROUP

3. REPORT TITLE

Proceedings of the Conference on Computer-Related Semantic Analysis.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Proceedings of Conference supported by National Science Foundation Grant, Office of Naval Research, & Air Force Funds.

5. AUTHOR(S) (First name, middle initial, last name)

Harry H. Josselson

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1966	330	

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
NSF - GN465 b. PROJECT NO. *Nonr 2562 (00)* c. *NR 049-128* d.	
	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Qualified requesters may obtain copies of the Proceedings from:
Slavic Department - Research in Machine Translation
Wayne State University, Detroit, Michigan 48202

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research, Wash., D.C US Air Force

13. ABSTRACT

A meeting dealing with problems pertaining to computer-related
semantic analysis was held in Las Vegas, Nevada, December 3-5,
1965. Scholars with experience in semantic analysis and/or
computer processing of semantic data were invited to address
the meeting. In addition to a keynote address from the
President of the Association for Machine Translation and
Computational Linguistics, thirteen papers were presented and
discussed at the meeting. The texts of the address and the
papers as well as summaries of the discussions edited primarily
for style and elimination of repetitious material comprise the
Proceedings. The main objective of this report is the dissemi-
nation of information to interested groups and individuals who
were unable to participate in or attend the meetings, since for
working purposes, attendance at the conference was restricted
to representatives of federally sponsored MT groups. A list of
conference participants and the program of the meeting are also
included.

DD FORM 1473 (PAGE 1)
1 NOV 65
S/N 0101-807-6811

Unclassified
Security Classification

A-31465

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Semantics						
Computer						
Lexicology						
Disambiguation						
Analysis						
Classification						

DD FORM 1473 (BACK)
1 NOV 65