

the FINITE STRING

NEWSLETTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS 

VOLUME 14 - NUMBER 5

SEPTEMBER 1977

ASPECTS OF ENGLISH SENTENCE STRESS, by Susan F. Schmerling,
 Reviewed by Dwight Bolinger 2

ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY,
 Vol. 11, edited by Martha E. Williams,
 Reviewed by Ralph Weischedel 25

FACTORS INFLUENCING THE PLACEMENT OF ENGLISH ADVERBS IN
 RELATION TO AUXILIARIES, by Sven Jacobson,
 Reviewed by Ivar Tonisson 29

FONDAZIONE DALLE MOLLE: Bibliography 1977 45

New Journal: LINGVISTICÆ INVESTIGATIONES 46

New Journal: STUDIES IN LANGUAGE 48

ARTS & HUMANITIES CITATION INDEX 50

CURRENT BIBLIOGRAPHY 51

AMERICAN JOURNAL OF COMPUTATIONAL LINGUISTICS is published by the Association for Computational Linguistics.

EDITOR: David G. Hays, 5048 Lake Shore Road, Hamburg, New York 14075.

EDITORIAL ASSISTANT: William Benzon

MANAGING EDITOR: Donald E. Walker, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California 94025

TECHNICAL ADVISOR: Martin Kay, Xerox Palo Alto Research Center

Copyright © 1977

Association for Computational Linguistics

A S P E C T S O F E N G L I S H
S E N T E N C E S T R E S S

SUSAN F. SCHMERLING

Department of Linguistics
University of Texas
Austin 78712

UNIVERSITY OF TEXAS PRESS
Austin 78712

xiv + 127 pp
\$8.95

1976

LC 76-6561
ISBN 0-292-70312-0

REVIEWED BY DWIGHT BOLINGER

2718 Ramona Street
Palo.Alto, California 94306

The bulk of the work here reviewed is a survey and criticism of previous studies of sentence accent in English, most of it recent. The critical portion of the book is its best feature. While there are constructive suggestions scattered throughout, only one brief chapter (ch. 5, 19 pages) is fully devoted to developing the author's views.

I modify S's terminology slightly to avoid what has been a persistent confusion. In place of word stress I use stress, and in place of sentence stress I use accent. Stress is a potential: it is an abstract entity which marks the syllable in a word that will be made prominent if the word is highlighted in an utterance. The

highlighting is done by accent, which is a contrast marked chiefly by a turn of pitch (not primarily, as used to be supposed, by loudness). Stress is thus a potential for accent. That it should be necessary to insist again and again on this principled distinction testifies to the difficulty of getting people to hear what goes on in intonation, and to abandon the terms they have grown used to, however misleading. One other term needs to be defined: A accent. This is the commonest of the pitch turns, consisting in a rapid downmotion either within the syllable or immediately afterward:

Jo It's hn.	John It's ny.
-----------------------	-------------------------

An A-rise (the term is D. R. Ladd's) has a rise after the fall:

hope I so	wo He uld!?!
---------------------	------------------------

Sentence accent is the accent with (presumed) maximum prominence in a sentence.

The main burden of the critical chapters is the N(uclear) S(tress) R(ule) of Chomsky-Halle 1968--a scheme applied to transformational cycles whereby the sentence accent is determined by the syntax: given the words and the syntactic structure, "the choice of stress [= accent] contour is not a matter subject to further independent decision" (Chomsky-Halle 1968, 25-6). S agrees³ with other critics that there is no such predetermination. The NSR not only fails to predict many instances of the location of the accent, but has to be patched up to work as well as it does.

Revisions of the NSR (ch. 3) are hardly more successful than the original. Bresnan 1971 adds an "ordering hypothesis" whereby

the accents of simple sentences "are preserved through derivations" (30). In the derivation of a sentence like I have instructions to leave there is an embedded direct object to which the NSR assigns a primary accent (leave instructions), automatically reducing the accent on leave. This direct object is then deleted, and the derivation passes to the next cycle, on which a primary accent is assigned to the instructions in the matrix sentence. S cites derivations in which Bresnan's hypothesis leads to wrong predictions.

Lakoff 1972 accepts the essential correctness of the NSR but revises it to apply at the level of surface structure. For various reasons, including both the inherited difficulties of the NSR itself and problems in interpreting the global rule that Lakoff proposes, S rejects this revision as well.

The critique then passes to Bolinger 1958, 1972. He denies the dependence of accent on syntactic structure, claiming that it reflects the speaker's intention to mark points of information focus. S for the most part accepts the negative side of **this claim**, but cites a number of examples that supposedly disprove the idea that semantic weight and relative predictability are the things that count. As the same examples are re-used in the chapter that states S's own theory, I reserve discussion of them till later.

The middle chapters (3, 4) examine certain assumptions that most studies of accent have made. Ch. 3 argues that there is no such thing as a neutral sentence with a "normal" position for the accent, and offers the interesting explanation that what the "normal accent" advocates have been referring to, without realizing it, is the citation form of a sentence--that this is the only truly "context-free" form that a sentence can have. Citation forms are

come,
Come,
come.

would not strike us as strange. I suggest that this tendency does manifest something that is "neutral" as regards highlighting any particular part of the sentence (see also Ladd 1977 for a similar idea), though it is not neutral in its own sphere: namely, the performative function of intonation, signifying something like 'I have this important thing to say' (more specifically, in a declarative sentence, 'I declare that'). The fact that the accent tends to come close to the end is a linguistic universal, having the psychological correlate of climax— a sentence has more impact when the most recently-spoken part of it is made prominent. The phenomenon is found in the most diverse languages. That it is not neutral in its own sphere, but contrasts with earlier positions of the accent, can be shown by such types as

God, they were all over the place! You couldn't escape them!

God, they were all over the place! You couldn't escape them!

There is no highlighting of over versus place here, but rather two levels of impact--the second, with its earlier accent, "holds back" Performative accent may occur on every word--even every syllable--of an utterance for maximum impact: I will not open the door! In I did it with my own two hands the word two serves little purpose except as the carrier of an extra accent. But the favored location is at the end, given the principle of recency. (Actually the next-to-last syllable, to accommodate the intonation turn--see Hyman 1975 and Bolinger 1977.)

If this is true, every sentence represents an adjustment be-

tween performative accent and accent on particular constituents. If besides putting the main accent at the end for maximum impact we also find a tendency to maneuver the least predictable element (the "comment" in S's scheme) to the end, it is probably to enable the accent to kill two birds with one stone.

S realizes that there is some "general phenomenon" (54) whereby the accent goes toward the end, but she gives only a negative characterization--if a string is "unanalyzable" it tends to have a terminal accent. An example is for Pete's sake, analyzable, versus for Pete's sake, unanalyzable exclamation (really the exclamation has the multiple accents noted above, with Pete's accented as well; it might otherwise seem that sake was being analyzed out for some purpose). But such idioms may move the accent for different degrees of impact when doing so will not appear to break them up (if indeed not in all cases--but the point is not worth arguing here). Give a damn 'care' is an instance:

I keep after them but they won't help. Those people just
don't give a damn.

... Those people just don't give a damn.

The latter censure is milder--it might imply 'What' the use of complaining?'

In parallel constructions the climactic accent is cumulative
from sentence to sentence:

He studied law. He studied medicine. He studied architec-
ture. He studied PHYSICS. But the essence of things
eluded him.

So S is probably right when she says that there is no such thing as a neutral context and no such thing as a "normal" accent,

but there does seem to be an accent that is not directly relevant to the semantic relations within the sentence. (In fact, more than one, but there is no space to elaborate here.)

The assumptions criticized in ch. 4 are two: that membership in a certain category (say noun or pronoun) may tell whether something is accentable, and that there is a special kind of accent that may be called contrastive.

S points out that there are so many exceptions to the reputed accenting of nouns and non-accenting of pronouns--even when one or the other is anaphoric--that no generalizations can be made. She cites the type I know who's standing in front of Mary, but I don't know who Mary's (she's) in front of to show the accenting of an anaphoric nominal. Later she explains this accent as the expression of a change in syntactic relationships (72). (She might have added that the change is sometimes made explicit in the syntax, with a shift of accent: but I don't know who Mary [she] herself is in front of. This is like the factoring out of tense from a verb: Why don't you write it? -- I wrote it! == I did write it!)

Though S is undoubtedly right in her judgments about the classes of nouns and pronouns in general, it is still worth while to look for subclasses that may be inherently unstressed. Possibly English has a loose set of "classifier" nouns such as are found in many languages. Why do we say 25-cent piece but 25-cent coin, The orange had a soft spot but The orange had a soft blemish for Pete's sake but for Pete's welfare Elm Street but Elm Lane?

As for "contrastive" accent, S argues that it cannot be characterized either phonetically or semantically. Her judgments are

surely correct. As a semantic explanation, contrastive is too specific. In place of it, S presents a looser generalization for which she uses such informal terms as that which is remarkable, noteworthy (67), or new (72), on the positive side, and lack of significance (73) or taken for granted (75) on the negative side. She embodies this opposition in her "Principle I: Certain items in an utterance are treated by the speaker as relatively 'insignificant' and fail to be assigned stress [= accent]" (75). She is careful to distinguish her "taken for granted" principle from the notion of presupposition (77-8). In a sentence like I didn't take my umbrella because I didn't realize it was raining the factive verb realize entails the factuality of 'raining'. But this has to do with truth, not with knowledge which the speaker and hearer share--there is no reason to take it for granted that the hearer knows it was raining at that place and time.

We may now look back at the criticism leveled at Bolinger for using the term predictability. Although this occurs in a context of several other informal designations comparable to S's own informal noteworthiness, such as information focus and semantic weight, S unfortunately picks up a statement that invites a narrower interpretation (40-42). The relevant example is

I don't care how many passengers were rescued from the Titanic. At that point in history what I want to know is what kings abdicated.

Abdicated is de-accented because "The speaker is interested in the fate of kings and the rise of democracies. There is enough mutual understanding between him and his interlocutor to make him reasonably sure that the mention of 'kings'--in the context of de-

mocracy--will suggest 'abdication'" (Bolinger 1972, 635). The passage in which this occurs began with the statement that "what counts is relative semantic weight" (with emphasis as shown), and the only claim intended was that as between 'kings' and 'abdicate', 'kings' had more to say--the question was 'kings' (~~and~~ 'monarchy') versus 'democracy', not 'abdicate' versus 'remain'. In S's terms, which I conceive to be naming the same thing that I was trying to name, 'kings' would be the element designating that which was most "remarkable".

S. continues her critique with two examples from her own experience, to disprove the idea that when the accent falls on the subject and not the verb, the verb has to be "predictable": (1) Truman died; (2) Johnson died. The first was spoken when Truman's ultimately fatal illness was on everyone's mind. The second came at a time when Johnson's health "was not on people's minds as Truman's had been, and when his death came it was a surprise". As S puts it, "Truman's death was expected, Johnson's was not. Bolinger's theory would appear to suggest, however, that the mention of Truman in the relevant context should have suggested 'death'--and accordingly died should not be accented. "On the other hand, the mention of Johnson in the relevant context should not have suggested 'death' any more than anything else one might have wanted to say about him", and therefore died should have been accented.

The error here is to assume the mention of Truman, making it part of the context. It is in the act of mentioning that the speaker has to decide which element is more suggestive and which is less.

To say that "Truman's death was expected" is to attach the expectation to the whole event, when it needs to be attached differentially to Truman or to death. The question is, given the total context, which word carries greater semantic weight, is more newsworthy, more remarkable--Truman or died? Everybody has been talking about Truman so an additional mention of Truman is not particularly newsworthy. On the other hand, the question at the time was, will he live through the day? will he linger on? will he perhaps rally? S's same misconception regarding what the "expectation" attaches to crops up later in her "topic-comment" cases. She cites the example John survived as one that "could be uttered only in a context where some such expectation on the part of the audience was assumed by the speaker (say, if we know that our friend John has been involved in a serious automobile accident)" (90). Expectation of what? Not that John would survive, but that John would be mentioned. If we know that he has been involved in a serious accident we no more expect him to survive than to die or to be knocked senseless or to come out without a scratch--which possibility will be the actual outcome is precisely the question. Another pair that points up the contrast more sharply:

What do you have against smoking? -- It's the nicotine that
it puts in your body.

... It's the harm that it does to your body.

Given the knowledge of the world that speaker and hearer share, which noun "tells more"? What we know or should know of nicotine clues us in: the speaker could have said It's the nicotine and let it go at that. But in the second sentence, harm is not particu-

larly informative--the question itself, in its have against part, could easily suggest 'harm'. The point now is 'body', that is, 'bodily health'. The speaker could have fronted body: It's your body, the harm that smoking does to it.

The same fronting is possible with S's examples:

What's the news today? -- He died. Truman, you understand.

What's the news today? -- Johnson. He died.

And so with S's additional examples:

Hey, your coat's on fire! = Hey, your coat! It's on fire!

Come on in--the door's open = Come on in--it's open.

(In the latter, it is not even necessary to mention door.)

It is not the newsworthiness of the utterance as a whole that counts, but the relative newsworthiness of the items to which relative pitch prominence is to be assigned. S's misunderstanding was compounded of my careless wording and her confusion between overall and differential expectation, or so it appears to me.

S adds some examples of a type where the noun appears to have little semantic weight of its own, yet carries the accent. They represent, I think, a balance between accent-for-prominence and performative or climactic accent. All are examples with be as main verb. I illustrate with examples of my own, to make the distinctions clearer:

1. Bears are a cantankerous species.
2. Bears are a species that is cantankerous.
3. Bears are a cantankerous lot.
4. *Bears are a lot that is cantankerous.
5. Bears are cantankerous.

Though everyone knows that bears are a species, this word is accented in both 1 and 2. The speaker is distinguishing among species, not merely describing bears--that is accomplished by 3 and 5; this accounts for the accenting of species as well as cantankerous. But why is it "more" accented in 1? The fact is that it is not, or need not be. "More" is an unexamined concept, prompted by the failure to treat performative accent as a separate entity. If species is more accented than cantankerous in 1, it is because of gradient climax. It may easily be less accented, that is, lower in pitch and intensity:

Bears^{RS}
are a can
t^{ankerous}spe
cies_s.

The observations up to this point have involved the contrast between accent and non-accent. The problem of relative pitch height brings a new dimension (see below). S's examples are neither proof nor disproof of the notion of predictability, newsworthiness, focus, or whatever one wants to call it. The very marking represents only an optional sequence--either accent can be "more" than the other--and the marking in many cases is simply false: the second of the two is a non-accent or a de-accent. The circumflex is a relic of Trager-Smith phonology, a confusion of accent with stress or vowel quality.

S's own proposals are mostly contained in ch. 6, where she presents three more principles (II, III, IV) plus a "rhythm rule". Principle III is merely a statement of the "rightmost-the-loudest" notion, which I have claimed to be true at least part of the time to the extent that performative factors make it so. Principle II

reads: "The verb receives lower stress [= accent] than the subject and the direct object, if there is one; in other words, predicates receive lower stress [= accent] than their arguments, irrespective of their linear position in surface structure" (82). This concerns "simple 'news sentences'", where "the speaker assumes no particular expectations with regard to the information content on the part of his audience" (81). Examples: Jóhn hít Bíll; Jóhn díed. The principle does not apply to sentences like John broke it or Jesus wept, where John and Jesus have already been introduced.

The main problem is that "news sentences" are not sufficiently defined. Take S's principle in the narrowest sense, where news means what McCawley has called "hot news"--something being mentioned for the first time:

1. What do you have against Smithers? -- Oh, he's not a bad chap personally, but his ideas make no sense.
2. How were the enemy able to infiltrate? -- All the s'entries had deserted their posts.
3. What made the train late? -- A ców had 'blocked the tracks.
4. What happened today? -- Marie Antoinétté just had her head chopped off.
5. Why are you looking so worried, Major? -- A Messerschmitt Two-Six-Twó just penetrated our defenses.
6. How come you were late for work? -- The 8:15 bús didn't put in an appearance.

(Even if "no particular expectations" means "discourse initial", some of these are still normal. Trainman says Listen, everybódy:

a cow's blocked the tracks; we'll have to change trains. Sansculotte says Hey, guess what! Marie Antoinette just had her head chopped off.) Do the direct objects in these sentences qualify as "arguments"? Do the expectations that justify the de-accenting of the objects qualify as "particular expectations"? In 1 there can be no specific expectation of the concept 'make no sense', though a listener will certainly expect ideas to communicate more. In 6, put in an appearance is an existential expression (see below)--the speaker could have said didn't show up, with no direct object; of course if a bus was "the trouble", one expects something like "not showing up" or "being late". As for 2 and 3, we naturally expect sentries to have posts and trains to have tracks, though if "information content" in S's definition means the event itself, nobody expected the posts to be deserted or the tracks to be blocked. As for 4 and 5, we can assume that 4 was uttered at a time when head-chopping was a national diversion and hence to be expected, and 5 in wartime when defenses were a daily preoccupation. And there is no real difference between cases like these and others where the expectations are more "particular", as in Who can solve the mystery? -- I think that John holds the key, with holds the key implicit in 'He can solve the mystery'. But consider S's prime example John died--isn't it the ubiquity of death that makes it possible for us to say that? Suppose he had not simply died, but exploded. Without some extremely heavy contextualization it would be impossible to say John explôded. The sentence would have to be divided--John--he exploded--in answer to a question like Why are you looking so glum? It is considerations like these--including just the usualness of some things--that cause speakers to place

emphasis as they do.

So even where the strongest case can be made--with "hot news"--S's principle leads back to relative semantic weight, to what does and what does not contribute most to making the point that the speaker wishes to make. Actually S does not mean "news" in this extreme sense of newness, because among her examples is the proverb Great oaks from little acorns grow, cited to show that the principle works regardless of the order, the accents being the same as in Great oaks grow from little acorns (83). As far as I can see, S has not considered proverbs with intransitive verbs, where, unlike the John died case, it is the verb that normally carries the final accent:

Power corrupts.

Old soldiers never die; they just

Frogs croak.

fâde away.

Time flies

(One can even find similar stereotypes referring to hot news: Aren't you staying? -- 'Fraid not; dûty calls.) S's proverb example is worth a second look. Take a pair such as

Mân his time must bide.

Mân must bide his time.

where, by S's accounting, bide and time would have to be said to get the main accent by virtue of their final position. In my view, the important thing is that both are accented in both sentences, and there is no reason to de-accent either one. In S's example there is a reason for de-accenting the verb: grow is a kind of existential--it is of relatively low semantic value, contributes little to the content of the sentence. Great oaks from little acorns sums it up. Much the same is true of the footnoted example (117)

One swallow does not a summer make--again, One swallow, hardly a summer says it all.

Another consideration is the nature of the subject. Even with hot news, indefinites may well not carry the main accent: Hey, mom! Something's burning!; Hey, mom! Some guy is screaming out here!

Still another is deixis. Hot news may readily accent the verb as well as the noun, where both are freshly introduced on the scene: Look! That pôor little bôy is crying! Look! A bunch of wômen are picketing! Look! That policeman is beckoning to you! The last example suggests that you had better heed the signal--if it were said Look! That policeman is beckoning to you!, with the verb de-accented, the speaker would be implying something like 'Isn't that interesting!'

But the most serious problem has to do with the notion of "levels". By trying to treat accent out of its intonational context, S leaves cases like the following undefined, where clearly hot news is involved (the speaker is reporting something he has just read):

Sa
 y, did you see this? Some es^{caped} pris^{oners} have mur^{dered} a man.

The "stresses" would be marked, I suppose, like this: Some escaped prisoners have murdered a man; but the important thing is the nature of the accents on prisoners (A-rise) and murdered (plain A). For there to be an "expectation", the usual thing would be a continuation of the rise on prisoners:

mur

...prisoners^{rs} have dered...

It is impossible to sort this out using S's discussion. One might suppose, first, that it is a case of two "intonational units" and that an acute accent should appear on both prisoners and murdered; but "intonational units" are defined as "phonological phrases" or "breath groups" (11)- and there is no evidence, aside from the pattern of the A-rise itself, for any separation between prisoners and have. (And S's discussion of her example Now Jôhn I like makes it fairly clear that she would not put a separation just on the strength of an A-rise. From the fact that she equates this example with I like Jôhn, which has A-rises on both like and John, I take it that she intends an A-rise on John in the first example- manifested minimally by a drop in pitch on I. Without it, the kinship is less-close.) Or one might suppose that this is one of S's topic-comment sentences, where the rule is to accent both the topic and the comment (Principle IV, 94); but then topic is defined as "something the speaker can assume to be, in a sense, on the addressee's mind, or immediately inferable from the total context"-- which can hardly be the case with some escaped prisoners. The fact is that separate items can convey news separately when each is given an A accent. In this sentence, both the presence of the prisoners on the scene, and the murder, are news.

The "topic-comment" notion is true by and large, but again suffers by being divorced from intonation. What can be "assumed to be on the addressee's mind" may take either of two forms. In one, the topic is 100% resumptive--it adds no more information than

would be contained in a de-accented pronoun. Typically there is a repetition, either literal or by sense:

I couldn't get in. The front window was closed. -- You're wrong. The front window was open.

Jack, Marion, and Lucy all think it's OK. -- All I can say is, your friends are crazy.

In the first, the front window is a literal repetition, and contains no pitch prominence whatever. In the second, the speaker can assume that the hearer will know that your friends is the same as Jack, Marion, and Lucy, and again there need be no pitch prominence. The trouble is that since pitch height is ignored, the "stress" markings will be the same, for instance on the front window, as if an accent were actually there.

The second kind of assumption takes some familiarity for granted, but still finds it desirable to single out the topic. This is done by accenting it, but with a lower pitch than the one on the comment:

What were we to do? Our gold had been confiscated. Our jewels had been stolen. Our clothes were missing. We were alone in a strange land.

That the speaker had had gold and jewels is assumed to be known (the noun phrases are definite), but they are too important to the situation to be de-accented. But this passage illustrates other things as well. For one, the "news" type Our clothes were missing occurs in a context of "topic-comment" types, yet there is no sensation of incongruity. It is only by insisting on a dichotomy between "news" and "topic-comment" that any inconsistency is created;

a theory that embraces both as examples of relative information value (relative importance, relative newsworthiness) causes no such trouble. The reason why missing is de-accented is that it is the least important of the three verbs in that series--it is an existential verb, like grow and make above, only more so: it is a lit-
eral existential, referring, as part of its meaning, to what is or is not on the scene. (What is brought on the scene--the entity introduced--is foregrounded even when expressed with an indefinite such as something or a noun such as thing. In

Something very fun

ny happened yesterday.

the normal thing is to put everything at high pitch up to the A drop. Similarly

A funny thi

ng happened yesterday.

The high plateau can be tilted either way, with affective differences--if tilted up it could express 'heightened suspicion'. But the important point here is that happened is an existential verb for events, and illustrates the downplaying of such verbs.) Another point here is the relative noteworthiness of confiscated and stolen. The first is the more unusual act (or would be except in a context like that of customs inspection), and to de-accent the word in this setting would be less likely. On the other hand, if Our jewels had been stolen preceded Our gold had been confiscated, stolen might readily be de-accented. There is no prior sentence with a parallel construction and an accented verb to serve as model

and foil. Stealing is not such a noteworthy act that jewels cannot outweigh it. Finally, the last sentence contains three accents no one of which is less prominent than the others--an example of something that S's treatment does not touch.

I conclude by taking three passages from S and testing them for adequacy. The question is whether news and topic-comment works better, or relative semantic weight.

The first involves the example John survived, which S says "seems impossible as an out-of-the-blue report" (90). The problem is to find a context in which survived carries relatively little information. If Noah had had a son named John, and John had finally made it to the ark along with some other stragglers, Noah might have exclaimed Jóhn survived! Thank God! Given the flood, survival was on everyone's mind, and no particular attention needed to be called to it--it would almost have been enough to say John! Thank God!

The second involves the example What's wrong with John? -- His dog was run over, classed as a "'news' example" (97). S's explanation is in terms of grammatical classes (subject versus verb) or logical form (arguments and predicates): the verb has "lower stress" by Principle II. But why should a speaker be governed by such a rule, which has no direct bearing on his intentions? Suppose we give minimal answers to What's wrong with John? as a way to discover what is most essential:

What's wrong with John? -- His dog. (Run over, you know.)

What's wrong with John? -- A hit-and-run. (His dog, you know.)

The second would be a strange answer, not because of the form, but because the wrong thing is highlighted. (Nouns usually code more information than verbs--a lexical count will show that there are vastly more nouns than verbs, revealing a tendency for nouns to be more precise and verbs more inclusive. But that is only a statistic.)

The third involves the example I didn't want to go because my hair was a mess. This is supposed to illustrate "news sentences in which the predicate has the form of a noun phrase" (94-5).

Mess is de-accented because it is "not very 'nouny'". True, it is not--it is what I have elsewhere termed a predicative degree noun, one with the characteristics of an adjective. But that is not why it is de-accented. If we choose a not-very-nouny word that has more impact than mess, it will be harder to de-accent: I didn't want to go because my hair was a fright. On the other hand, by adding such a, which normally applies to something known, we can get away with de-accenting either mess or fright: I didn't want to go because my hair was such a mess (such a fright).

S has done a fine job of exposing the shortcomings of grammatical approaches to accent, and where her own work fails it is mainly because of a last vestige of trying to put things in grammatical or logical terms: nouns versus verbs, arguments versus predicates. Her training in syntactic theory, plus a keen ear for prosodic contrasts, make her one of the few persons who could challenge the current theories on their own grounds. One can only regret that so much energy and fine talent had to be spent on disproving an elaborate set of fallacies. A lot that goes on

in linguistics reminds one of the Gross National Product, which adds together the sums spent on causing cancer and the sums spent on preventing it. It would be nice if we could be constructive more of the time.

References

- Bolinger, Dwight. 1958. A theory of pitch accent in English.
Word 14.109-149.
- _____ 1972. Accent is predictable (if you're a mind reader).
Language 48.633-44.
- _____ 1977. Intonation across languages. (to appear)
- Bresnan, J. W. 1971. Sentence stress and syntactic transformations.
Language 47.257-80.
- Chomsky, A. N., and Halle, Morris. 1968. The sound pattern of
English. New York: Harper and Row.
- Hyman, Larry. 1975. On the nature of linguistic stress. Preprint.
- Ladd, D. R. 1977. Light and shadow: a study of the syntax and se-
mantics of sentence accent in English. Preprint.
- Lakoff, G. W. 1972. The global nature of the nuclear stress rule.
Language 48.285-303.

ANNUAL REVIEW OF
INFORMATION SCIENCE AND TECHNOLOGY
VOLUME 11

MARTHA E. WILLIAMS, EDITOR

AMERICAN SOCIETY FOR INFORMATION SCIENCE

P.O. Box 19448 Washington, D.C. 20036

457 pages
\$28 members
\$35 nonmembers

1976

LC 66-25096
ISBN 0-87715-212-8
ISSN 0066-4200
CODEN ARISBc

REVIEWED BY RALPH WEISCHEDEL

Department of Computer Science
University of Delaware, Newark

Three chapters at the interface of computer science, information science, and library science are considered here. Our greatest attention is focused on the chapter on natural language processing in the Annual Review of Information Science and Technology (ARIST)

Volume five of ARIST, published in 1970, contained a tutorial by Harry Huskey on computer technology. In the present volume, the chapter "Computer Technology - An Update" by Philip L. Long provides a brief tutorial on significant hardware advances

since the 1970 review. The tutorial is primarily addressed to information scientists not familiar with recent hardware technology. Its purpose is to recount the advances in off-the-shelf components, such as disk packs, interactive terminals, and microprocessors, particularly as they affect large online information systems. This purpose is effectively achieved by the level of detail selected for explanation; the shape of the technological advances is presented without inundating a computer novice with too much technical detail.

"Online Systems, - Techniques and Services" is written by Beatrice Marron and Dennis Fife. Bibliographic online systems receive the most attention, since they are predominant in the literature. The review concentrates throughout on studies and analyses of the impact, trends, problems, and future of online services; therefore, it should be of value to all interested in online services. The chapter is well-written throughout.

Though a review of the interface of computational linguistics, artificial intelligence, and information retrieval is beyond the scope of the chapter, several aspects of the chapter will be interesting to those of us active in artificial intelligence and computational linguistics. The authors note that the areas of "natural language, semantics, inference and deduction, information organization and association" have progressed slowly, though these problem areas "are critical if online information systems are to become everyday tools in general problem solving and research." (p.166).

Also, the section on the user interface should be a valuable source of references for anyone conducting research in artificial intelligence approaches to data base user interfaces. The references cited there discuss criteria needed for a good interface, problems of providing such, advantages and disadvantages of requiring a human intermediary for occasional users, and issues in user training.

"Automated Language Processing" by Fred J. Damerau reviews research trends in natural language processing during 1974 and 1975, though some references from 1973 are included as well. To keep the review manageable, only articles published in English are included; also, work in which language is treated as uninterpreted character strings is excluded. Contributions from artificial intelligence, cognitive psychology, and linguistics are reviewed.

The author states (p.108) "The aim of the entire review is to guide an interested reader to the most significant or widely read literature, while making him at least aware of the difficulties and problem areas which may not be emphasized in the works cited." Personally, I find that to be the most valuable purpose such a review can serve. He admirably achieves this purpose, and his treatment of the topics is thorough.

For instance, in discussing semantic nets, he first points out their widespread acceptance as a representation of knowledge. The review continues, mentioning some of the theoretical problems with semantic nets, such as representing both extensional and in-

tensional descriptions of objects and representing quantifiers. Then, proposed solutions and the articles containing them are described. As an aid to the uninitiated, several articles illustrating the use of semantic networks in concrete language processing systems are mentioned. In addition to pointing out some technical weaknesses, the author also draws attention to a weakness in our methodology of reporting on systems. On pages 110-111, he states, "It is hoped that the comments of Woods and others on the deficiencies and limitations of the past uses of networks will be remembered in future implementations. Precise specification of the characteristics and limitations of each component in an understanding system is necessary in order to evaluate meaningfully one piece of work against another."

Because of the stated aim of the author, the chapter will be very valuable to those wishing to learn of the field. However, it is also valuable for those of us thoroughly familiar with the field, to have a review of the weak spots in the work and in our approaches.

FACTORS INFLUENCING THE PLACEMENT
OF ENGLISH ADVERBS IN RELATION TO
AUXILIARIES: A STUDY IN VARIATION

SVEN JACOBSON

ALMQVIST & WIKSELL INTERNATIONAL

P.O. Box 62
S 101 20 Stockholm - Sweden

599 pages

1975

SKr 106

REVIEWED BY IVAR TÖNISSON

Postdoctoral Scholar
Department of Philosophy
Stanford University
Palo Alto, California 94305

*The surface word order of adverbs in English is worth serious study, since the syntactic rules that are needed to describe it must be capable of capturing quite intricate syntactic relations. These rules would in turn give us new insights about the way natural languages work. However the syntax of adverbs is still not adequately understood, precisely because of the complexity of the phenomena involved. The standard reference work in this area remains still to be written, although several attempts have been made to tackle some of the problems in this field. Unfortunately most of

these are quite controversial, and little agreement exists between the various authors.⁽¹⁾ Hence Sven Jacobson's book is a contribution to a part of this field of research.

Jacobson tries to give a comprehensive description of preverbal adverb placement in surface structure. Following Jacobson I will refer to preverbal adverbs as preverbs. The term "preverb" was coined by Robert Lees. He introduced preverbs by means of the following phrase structure rule (where MV stands for the finite verb and its modifiers, objects, etc.).⁽²⁾

$$(1) \quad VP \longrightarrow (\text{Prev}) \text{ Aux} + \text{MV}$$

He also remarked that their normal (surface) position is after the second member of the auxiliary in most sentence-types.⁽³⁾ Owen Thomas used the same phrase structure rule to introduce the notion of "preverb", but he argued that they occur normally after the first auxiliary verb.⁽⁴⁾ Hence, according to Lees and Thomas, preverbs are non-sentential adverbs, since they are directly dominated by the node VP and not by the node S. Edward Klima, however, introduced preverbs in a totally different way.⁽⁵⁾ All the items that he included in the category Adu seem, in fact, to be preverbs. Hence, his way of introducing the notion of preverb maybe represented by:

$$(2) \quad S \longrightarrow (\text{Wh}) (\text{Neg}) (\text{Prev}(\text{Neg})) (\text{Prev}) \text{ Nominal} - \text{Predicate}$$

Clearly he regards preverbs as sentence adverbs since they are directly dominated by the node S. Charles Fillmore takes a similar approach. He introduced the notion of "preverbs" by the phrase

structure rule (3):

$$(3) \quad S \longrightarrow (Q) \text{ (Prev) NP + Aux + VP}^{(6)}$$

Hence, he too regards preverbs as sentence adverbs.

In contrast to the authors cited before Barbara Partee claims that "preverb" is not a syntactic category but rather a feature shared by all sentence adverbs.⁽⁷⁾

Jacobson, in an earlier work, defines preverbs as adverbs in surface structure that occur between NP and V, or between NP and adjectives in predicative syntactic position.⁽⁸⁾ He adds that in phrase-markers preverbs can be immediately dominated either by the node S or by the node VP and, hence can be either sentence adverbs or verb-phrase adverbs respectively.

In the book under review regards "preverb" as an entirely positional notion and defines it as follows:

(I) A verb phrase adverb that precedes a verb node, if the adverb and the verb in question are immediately dominated by V in the surface structure. Evidently he assumes that the V node can sometimes be rewritten by phrase structure rules as Verb i.e., (4) holds.

$$(4) \quad V \longrightarrow \cdot \cdot \text{ " } \cdot \text{ Verb } \cdot \cdot \cdot \cdot$$

(II) A sentence adverb ~~that precedes~~ a V node, if the adverb and V node are immediately dominated by the S node.⁽⁹⁾

In contrast to his earlier definition, adjectives in predicative position are assumed to be verbs, and hence do not need special mention. On this point he agrees with the Generative

Semanticist. I believe, however, that this is an untenable syntactic view which has rightly been criticized.⁽¹⁰⁾ Hence, I find his failure to mention adjectives in predicative position unacceptable.

Now it is clear that his idea of an entirely positional notion presupposes a theoretical framework within which the relative position of preverbs in the surface structure can be discussed, and this framework can of course be questioned. Hence Jacobson's notion of position is not as expected a purely empirical one, contrary to what one might suppose when one looks at his use of statistical data. Moreover the theoretical framework he uses has to be defined somewhere, which he unfortunately does not do. The reader should keep this in mind when I will discuss whether Jacobson has laid an adequate foundation for his quantitative study.

It is perhaps worth noting that Rodney Huddleston seems to agree with Jacobson's idea that "preverb" is an entirely positional term. For he claims that a "preverb" is so called because of its most usual position.⁽¹¹⁾

To conclude, there seems little agreement in the literature about how to define or even introduce the notion of "preverb." There is perhaps more agreement about which adverbs are actually preverbs, but so far the different authors have not given sufficiently comprehensive lists of preverbs to determine whether this is the case. It is not even clear that there exists a syntactically interesting group of adverbs called preverbs worth extended study

as a unit, Jacobson has not given any evidence that there is. For the time being, however, I will assume that it is worthwhile to consider "preverbs" as a group.

The book's aim is to present a survey of how preverbs are actually used by carrying out a quantitative study, but as I already mentioned this study depends also on the theoretical framework chosen by the author. His statistical analysis is richly illustrated by examples, consisting of actual usage of American English which he draws from a finite corpus of written sources. He thus succeeds in avoiding the marginal examples that many linguists use. However, there are serious problems connected with his approach, since a finite corpus seldom captures adequately all of a native speaker's intuitions about his language. (See for example Samuel Keyser's review of Jacobson's dissertation.⁽¹²⁾ It illustrates nicely the dangers of using one's own intuition about the language of which one is a native speaker. Jacobson points out on P. 13 that one of Keyser's examples is odd.⁽¹³⁾ He could strengthen his claim to say that Keyser's example "John will send the money back to the girl roughly" is semantically deviant. Hence Keyser has not been careful enough when using his intuition as a native speaker to illustrate his syntactic points. Still, it also shows the weakness of Jacobson's approach solely using a finite corpus.) Furthermore, one cannot be sure of finding the right types of examples, since the probability that a suitable example occurs in a given corpus can be very low even if the corpus is quite large. A better method, in my

opinion, would be to combine the use of a corpus with elicitation from theoretically unbiased informants. On pp. 12-15, Jacobson expresses some doubts about the value of elicitation, but I believe that he underestimates its value. Using elicitation, crucial sentences lacking from a corpus can be fabricated and tested in an unbiased way. "Uncontrolled" intuitions can be used to discover what crucial sentences need to be tested. Clearly this method is preferable to using the biased intuitions of the theory constructor to test his own theories. It is important to notice here that in the theory construction stage any guess - no matter how biased it is - is acceptable, if it leads one to make correct predictions, (but of course such a situation is unlikely to occur). However, in the theory testing stage one needs unbiased empirical data to check, if one's predictions are correct or not. This fact obtains for all sciences, and is just the way the hypothetico-deductive method works. Jacobson seems to be unaware of this, when, on p. 13, he talks vaguely about being both an empiricist and a rationalist. Empiricism and rationalism are epistemological theories that have no direct bearing on theory construction and theory testing.

To conclude, Jacobson should make a distinction between the theory construction stage and the theory testing stage of scientific activity, and leave room for hunches and biased intuitions as well as unbiased empirical data.

Transformational grammar, it should be pointed out, is still by and large in the theory construction stage so that any insight

of any nature is important, and not in the theory testing stage, where unbiased empirical data is essential. At present it is very easy to falsify any explicitly formulated grammar by considering examples that have not been understood by the theory constructor, and hence are not adequately accounted for by his theory. It is common knowledge that as yet no adequate grammar for any natural language has been constructed, and that everybody is a long way from finding one which can seriously be tested on the whole of a natural language.

Jacobson's attempt to gather quantitative data about some adverbs is in principle laudable, but I have reservations about how the attempt is actually carried out.

His book begins with a short presentation of the aims of the work. He finds the relation between preverbs and auxiliaries to be especially interesting. Hence to reiterate he intends to present a survey of how preverbs are actually used by carrying out a quantitative study, illustrated by examples. The syntactic aim of the book is to give a comprehensive description of preverb placement in surface structure; the taxonomic aim, to give the necessary discrete categories is subsidiary to this. In his study he finds that there is no significant difference between American and British English. The tests on which Jacobson bases his conclusions represent many different types of prose, spoken as well as written. All of them were produced in the post-war period primarily in the 1960's. The main part of the book consists of a detailed discussion of the corpus, about specific preverbs and a presentation of the

statistical methods used.

I do not propose to comment on Jacobson's use of statistics, since I am not qualified to do so. I now want to turn to a discussion of some details of Jacobson's work.

Jacobson remarks on p. 7 that the relations of preverbs to auxiliaries is especially interesting. It seems to me, however, that he fails to distinguish between stylistic and syntactic or semantic reasons for the non-occurrence of certain sentences. His account of this relation is therefore deficient. For neither stylistically inappropriate, nor syntactically or semantically deviant sentences do occur in Jacobson's type of corpus. (A clumsy sentence marks stylistically inappropriate uses, while a bad sentence marks syntactic or semantic deviance.) Hence he cannot tell whether a sentence is absent because it is deviant or because it is stylistically inappropriate. Elicitation tests, of course, could distinguish between the two cases. Maybe stylistic factors play no role here, but they can only be excluded by empirically based results.

Jacobson also has a problem with his discrete categories. On p. 9 and p. 49 he states that his taxonomy requires countable items and cannot therefore use the concept of a continuum. Hence Jacobson has to find a way to classify preverbs that gives rise to rather sharp boundaries, but I do not think that he has succeeded in doing this. This shows a weakness in his theoretical framework.

On p. 23 Jacobson characterizes what preverbs do syntactically and semantically. Syntactically verb phrase preverbs modify

the verb, and sentence preverbs modify the clause to which they belong to. Semantically verb phrase preverbs characterize processes or states; sentence adverbs characterize propositions (i.e., the semantic contents of the sentences or clauses), acts of communication, and events or circumstances. However, these functions are not unique to preverbs. This leads one to wonder why is the adverb position before the main verb should be interesting enough to warrant a special label. (If Barbara Partee is right it is even harder to motivate a special label.) What do preverbs have in common that is exclusive to them besides their position? If it is only their position, one could just as well study groups of adverbs called "presubjects" or "postverbs" too.

It should also be mentioned that Jacobson's characterization of the functions of verb phrase and sentence adverbs is inadequate.⁽¹⁴⁾

On p. 23 Jacobson says that a preverb can presuppose that the proposition expressed by the sentence the preverb occurs in states a fact. Now according to Enrique Delacruz, "presuppose" can have three distinct senses:

- 1) A sentence p may presuppose a sentences p*,
- 2) The speaker of a sentence p may presuppose a proposition p*,
- 3) A predicate or verb V may presuppose a sentence p*.

It is not clear to me which of these senses of "presuppose" Jacobson has in mind. Here Jacobson could avoid this unclarity.

I disagree with Jacobson's statement on p. 25 that the generation of a preverb is often hinted at by means of a paraphrase which is less remote from deep structure than the preverb itself. Ray Jackendoff has shown that this type of paraphrase does not work in general and hence, is of limited usefulness.⁽¹⁵⁾

On the same page Jacobson talks about an uttered clause being subordinate to some hypersentential clause; which is wholly or partially deleted in the course of the transformational deviation. He should have explained his point a little more, since it is far from clear what those deleted hypersentential clauses are. Similarly, his phrase "sentence preverbs that are derivable from sentences on different levels in the structural hierarchy" needs more explanation to be comprehensible.

Jacobson goes on to expound on pp. 25-27 what I consider to be a confused and erroneous view, namely that a preverb can simultaneously be both a sentence and verb phrase modifier. He explains this in terms of a notion of "coalescence" (i.e., double derivation). But coalescence simply indicates that the sentence in question is ambiguous. Consider, for example, the following example given by Jacobson: "They were suddenly attacked." One reading that (corresponding to the case where "suddenly" is a sentence adverb) is that there was an attack on them that was sudden. The other reading (corresponding to the case where "suddenly" is a verb phrase adverb) is that they were attacked in a sudden manner. These two readings have different truth-conditions, and the sentence is therefore

two-way ambiguous. The preverb is a sentence adverb on one reading and a verb phrase adverb on the other, but never simultaneously both.

Jacobson considers a manageable number of preverb classes on p. 49. His classification is based on syntactic and semantic criteria. Now it is relevant to ask why syntactic and semantic criteria taken together should delimit natural preverbal classes. It might be that some classes of preverbs are syntactically natural while others are semantically natural, but not both, and vice versa. After all, syntax has mainly to do with the distribution of morphemes that determines the well formedness conditions for a natural language, whereas semantics has mainly to do with the information contained in the sentences that accounts for which inferences are valid in a wide sense of the term "valid". Jacobson fails to show that his combined syntactic and semantic criteria do indeed delimit natural classes of preverbs. Once more Jacobson fails to motivate adequately his theoretical framework.

The distinctions used to introduce the twelve classes of preverbs on pp. 52-66 seem somewhat arbitrary. Why should one pick these among the large amount of other distinctions that have been proposed in the literature.⁽¹⁶⁾ Jacobson should at least have tried to motivate why one should choose his distinctions but he has not done that.

There is a general problem about how one can test empirically Jacobson's claims about the semantics of preverbs. Maybe some of

his distinctions are too subtle to be empirically tested, because they require Jacobsonian semantic intuitions. Here is a case where one wishes that Jacobson should have a more empirical orientation that would make his classification more suitable to actual empirical tests.⁽¹⁷⁾ It is true that he uses an interesting corpus as his empirical data, but his analysis of the corpus, rests heavily on his own semantic intuitions. This is another illustration of the theoretical inadequacy of his study. Consider, for example, his discussion of intra-clausal temporal preverbs on pp. 52-53. Intra-clausal temporal preverbs denote a large variety of temporal aspects, namely the following: point of time, e.g., today; duration, e.g., long, frequency, e.g., often, temporal proximity, e.g., soon, and preceding coinciding and subsequent time in relation to the moment of speaking or writing, e.g., before, now, and later. Now Jacobson claims that many temporal preverbs have homonyms expressing verbal manner. He considers the following two sentences: (a) ". . .an offer which has been accepted is immediately extinguished," (b) "Now the argument of this book is not immediately concerned with the truth or falsity of what we say about China or Russia." He argues that "immediately" is a intra-clausal temporal preverb in (a) but that it expresses verbal manner in (b). However "immediately" can be analysed as expressing some kind of proximity in both (a) and (b). In (a) it is temporal proximity, and in (b) it is proximity between the argument and the truth or falsity of what we say about China or Russia. On this analysis "immediately" in (a) and in (b) are not

homonymos. The other examples can be dealt with similarly. (18)
 This illustrates yet again the dependence of Jacobson's classification on a theoretical framework which is nowhere motivated or even articulated in his book.

On p. 64 Jacobson claims (without supporting evidence) that if a paraphrase of a sentence containing "almost", "It was almost the case that. . .", it is possible then this occurrence of "almost" is a sentence adverb. This claim is, however, far from self evident and, in fact, other tests for an adverb being a sentence adverb has been proposed in the literature that are different from this test, but some of the less successful tests proposed are rather similar to Jacobson's. (19) I think that Jacobson's type of test is not adequate and needs to be replaced by other types of considerations. (20)

Also on the same page Jacobson claims that it is clear that "almost" and "nearly" are sometimes sentence adverbs. In contrast many authors have assumed that "almost" and "nearly" are always non-sentential adverbs. (21) Of course given Jacobson's test for sentence adverbs "almost" and "nearly" are indeed sometimes sentence adverbs, but as I mentioned his test is not acceptable.

I believe that Jacobson's book is a praiseworthy attempt to investigate a special class of adverbs empirically, but the details of his study leave something to be desired.

FOOTNOTES

*I would like to thank Anthony Ungar for commenting on an earlier version of this article,

- (1) See for example Renate Bartsch, (1976). The Grammar of Adverbials. Amsterdam, North-Holland, Romane Clark, (1970). "Concerning the logic of Predicate Modifiers" in Nous, (pp. 311-335), Jonnie Geis, (1970). Some Aspects of Verb Phrase Adverbials in English. Unpublished dissertation, University of Illinois at Urbana-Champaign, Ray Jackendoff, (1972). Semantic Interpretation in Generative Grammar. Cambridge, Massachusetts. MIT Press, Don Nilsen, (1972). English Adverbials. The Hague. Mouton, Malcolm Rennie, (1974). Some Uses of Type Theory in the Analysis of Language. Canberra. Department of Philosophy, Research School of Social Sciences, The Australian National University. Mono-graph Series, No. 1, Barry Taylor. The Semantics of Adverbs, (1974). Unpublished dissertation. Oxford University, Richmond Thomson and Rober Stalnaker, (1973). "A Semantic Theory of Adverbs" in Linguistic Inquiry. Vol. IV, (pp. 195-220).
- (2) See Robert Lees, (1960). The Grammar of English Nominalizations. Bloomington. The Indiana University Research Center in Anthropology, Folklore and Linguistics, p. 5.
- (3) See Robert Lees, (1962). "The Gramatical Basis of Some Semantic Notions" in Monograph Series on Languages and Linguistics, No. 13. Georgetown. Georgetown University Press, (pp. 5-20), p. 13.
- (4) See Own Thomas, (1966). Transformational Grammar and the Teacher of English. New York. Holt, Rinehart and Winston.
- (5) See Edward Klima, (1964). "Negation in English" in The Structure of Language ed. by Jerry Fodor and Jerrold Katz. Englewood Cliffs, Prentice-Hall, (pp. 246-323), pp. 254, 260, 262, 316-318.
- (6) See Charles Fillmore, (1967). "On the Syntax of Preverbs" in Glossa, Vol. 1, (pp. 91-125), p. 104.
- (7) See Barbara Partee, (1973). "Negation" in The Major Syntactic Structures of English by Robert Stockwell, Paul Schachter, and Barbara Partee. New York. Holt, Rinehart and Winston, (pp. 230-293). p. 267.
- (8) See Sven Jacobson, (1971). Studies in English Transformational Grammar. Stockholm. Almqvist & Wiksell. p. 31.
- (9) See p. 20 of the book under review.

- (10) See for example Joan Bresnan, (1976). "On the Form and Functioning of Transformations in Linguistic Inquiry." Vol. 7, (pp. 3-40), and Peter Culicover, (1977). "An Invalid Evaluation Metric" in Linguistic Analysis. Vol. 3. (pp. 65-100).
- (11) See Rodney Huddleston, (1976). An Introduction to English Transformational Syntax. London. Longman, p. 87.
- (12) Many doubts remain, however, the notion of transportability, which Jacobson himself accepts, makes the claim that "preverb" is an interesting syntactic unit especially dubious. See Samuel Keyser, (1968). "Review of Adverbial Positions in English." by Sven Jacobson in Language. Vol. 44, (pp. 357-374). pp. 368ff, Jackendoff op. cit. pp. 67-68, 80-81, 95, 106 and Jacobson (1971) op. cit. p. 83. If preverbs can be moved around in a sentence, why should the preverbal position be considered as especially important? Jacobson fails to answer this question.
- (13) See Keyser op. cit. examples (33)a - (40)b. The problem is that you either send the money or not, but there is no way to send money roughly. However you can decide to do something that means roughly that you send back the money, but this reading is hard to get for the native speakers I have asked. This holds probably because word order in English suggests preferred readings but does not exclude completely the other now-preferred readings. In general it is bad strategy to base your illustration of theoretical points on language examples that are at best marginal.
- (14) See my unpublished Stanford dissertation A Formal Semantics for Adjectivals and Adverbials. (1976). pp. 23-51 for adequate ways to distinguish between sentence and non-sentential adverbs.
- (15) See Enrique Delacruz, (1976). "Factives and Proposition Level Constructions in Montague Grammar" in Montague Grammar, ed. by Barbara Partee, New York Academic Press, (pp. 177-199). p. 179.
- (16) See for example Yehoshua Bar-Hillel, Jonathan Malino, and Avishai Margalit. "On Logic and Theoretical Linguistics." (1975) in Current Trends in Linguistics, Vol. 12, ed. by Thomas Sebeok. The Hague, Mouton. (pp. 37-101). p. 76, Bartsch op. cit., Clark op. cit. Frank Heny. (1973). "Sentence and Predicate Modifiers in English." in Syntax and Semantics, Vol. 2 (pp. 217-245). New York. Academic Press, and Rennie op. cit. In my dissertation I give a semantic classification of adverbs that consist of eleven classes, and try to motivate why one should choose this specific classification.
- (17) See Chapter 3 of my dissertation.
- (18) To further support my claim consider Jacobson's treatment of "briefly." Again he claims the adverb is an intra-clausal temporal preverb in (a) but expresses verbal manner in (b).

- (18) (Con't) His sentences are: (a) ". . .that great void in his soul which bitterness and rebellion had briefly left vacant," (b) "Our preliminary remarks about the constitution of the United States may, then, be briefly summarized." In this case however "briefly" can be understood to express a short spatio-temporal event. In (a) the event is the short event when his soul is void, and in (b) the event is the short event during which the remarks can be spoken or read. Hence (a) and (b) do not contain a homonymous use of "briefly" as Jacobson claims. Once more an alternative analysis destroys the assumption about homonymity.
- (19) See Thomason and Stalmaker op. cit. Criterion 4, p. 205 for a test that is very similar to Jacobson's. For other test of whether an adverb is a sentence adverb see footnote 14.
- (20) See footnote 14.
- (21) See my unpublished article "A Semantics for "Almost"", presented at the conference on Montague Grammar at SUNY at Albany, April 24, 1977. In it I present a extensive bibliography of the work done on "almost".

B I B L I O G R A P H Y

WORKING PAPERS (FEBRUARY 1977)

FONDAZIONE DALLE MOLLE

INSTITUT POUR LES ETUDES SEMANTIQUES ET COGNITIVES

17, RUE DE CANDOLLE, CH-1205 GENEVE, SWITZERLAND.

18. Semantics, preference and inference: A full description of a system and a program. Margaret King, Yorick Wilks. Forthcoming.
19. Frames, planes and nets. Greg Scragg. 40 pp. S.Frs. 5.--
20. A structure for actions. Greg Scragg. 40 pp. S.Frs. 5.--
21. On understanding German noun clusters. Wolfgang Samlowski. 20 pp. S.Frs. 2.50
22. A brief on Case. Eugene Charniak. 30 pp. S.Frs. 2.50
23. A process to implement some word-sense disambiguators. Philip Hays. 36 pp. S.Frs. 2.50
24. On the referential attributive distinction. Eugene Charniak. 55 pp. S.Frs. 5.--
25. Several ways to be suggestive. Margaret King. 63 pp. S.Frs. 5.--
26. Pragmatic aspects of noun cluster understanding in German. Wolfgang Samlowski. 20 pp. S.Frs. 2.50
27. Un sistema colloquiale di domandarisposta in lingua italiana. Bernd Treusch, Bruna Zonta, Franco Lestuzzi. Out of print.
28. A framed PAINTING: The representation of a common sense knowledge fragment. Eugene Charniak. 60 pp. S.Frs. 5.--
29. How to connect frames. Manfred Wettler. 54 pp. S.Frs. 5.--
30. Mutatis Mutandis I. Margaret King. 29 pp. S.Frs. 2.50
31. Ms. Malaprop, A language comprehension program. Eugene Charniak. 22 pp. S.Frs. 2.50

Note: A listing of Working Papers 1 - 17 can be found in AJCL Microfiche 22: 46-47.

NEW JOURNAL:

L I N G V I S T I C Æ I N V E S T I G A T I O N E S —
INTERNATIONAL JOURNAL OF FRENCH LINGUISTICS
AND GENERAL LINGUISTICS

EDITORIAL POLICY: French studies; General Linguistics; Theoretical
and Descriptive studies; Transformational+Generative Grammar.
Articles in French and English.

SPONSORSHIP: Joint publication by the Linguistics Department of the
University of Paris-8-Vincennes and the Laboratoire d'Automa-
tique Documentaire et Linguistics of the C.N.R.S. (University
of Paris-7).

EDITORS: JEAN-CLAUDE CHEVALIER, Université de Paris-Vincennes
MAURICE GROSS, Université Paris 7
CHRISTIAN LECLERE, Laboratoire d'Automatique Documentaire
et Linguistique (C.N.R.S.)

PUBLISHER: John Benjamins B.V.
Amsteldijk 44
AMSTERDAM, The Netherlands

SUBSCRIPTION (c. 450 PAGES, 2 ISSUES):

Hfl. 120,-+postage: to Libraries and Institutions, upon
receipt of an invoice.

Hfl. 80,-(postage incl.): for private subscribers, order to
be placed directly with the publisher

COMITÉ DE PATRONAGE / CONSULTING EDITORS

M. BRUGUIERE, Haut Comité de la Langue Française
 N. CHOMSKY, Massachusetts Institute of Technology
 A. CULIOLI, Université Paris 7
 A. DUGAS, Université du Québec à Montréal
 O.-L. GRUNDT, Université de Bergen
 M. HALLE, Massachusetts Institute of Technology
 R. JAKOBSON, Université de Harvard
 F. KIEFER, Université de Budapest
 S. Y. KURODA, Université de Californie à San Diego
 A. LENTIN, Université René Descartes, Paris
 T. M. LIGHTNER, Université du Texas
 J. LYONS, Université d'Edimbourg
 T. de MAURO, Université de Rome
 Ch. ROHRER, Université de Stuttgart
 S. SCHANE, Université de Californie à San Diego
 M. P. SCHUTZENBERGER, Université Paris 7
 E. SPANG-HANSEN, Université de Copenhague
 O. SÖDERGÅRD, Université de Lund
 R.-L. WAGNER, Ecole Pratique des Hautes Etudes

COMITÉ DE LECTURE / EDITORIAL BOARD

J.-P. BOONS, Laboratoire d'Automatique Documentaire et Linguistique
 (C.N.R.S.)
 J. DUBOIS, Université de Paris-Nanterre
 G. FAUCONNIER, Laboratoire d'Automatique Documentaire et Linguistique
 (C.N.R.S.)
 A. GUILLET, Laboratoire d'Automatique Documentaire et Linguistique
 (C.N.R.S.)
 R. S. KAYNE, Université de Paris-Vincennes
 J.-C. MILNER, Université Paris 7
 N. RUWET, Université de Paris-Vincennes
 M. SALKOFF, Laboratoire d'Automatique Documentaire et Linguistique
 (C.N.R.S.)
 J. STEFANINI, Université de Provence
 J. R. VERGNAUD, Laboratoire d'Automatique Documentaire et Linguistique
 (C.N.R.S.)

CONTENTS OF NUMBER 1

Gilles FAUCONNIER, Polarité syntaxique et sémantique
 Maurice GROSS, Une analyse non-presuppositionnelle de l'extraction
 dans *C'est. . .Qu.*
 Sige-Yuki KURODA, Description of presuppositional phenomena from a
 non-presuppositional point of view
 Théodore M. LIGHTNER, A note on theoretical inadequacies
 Claude MULLER, A propos de *de partif*
 Alain ROUVERET, Les consécutives: forme et interprétation
 Morris SALKOFF, Les principes d'un analyseur syntaxique du français

NEW JOURNAL: S T U D I E S I N L A N G U A G E

EDITORIAL POLICY: Linguistics and Philosophy, the foundations of language; descriptive, formal, typological studies; interdisciplinary contributions - sociology, psychology, mathematics, AI, law, ethics, logic, phenomenology of language, etc.

EDITORIAL BOARD: John W.M. Verhaar, Managing Editor
University of Jakarta, Indonesia

Werner Abraham, Review Editor
University of Groningen, Netherlands

Richard D. Brecht
Harvard University, USA

Bruce Fraser
Boston University, USA

Morris Halle
M.I.T., USA

Peter Hartmann
Univ. Konstanz, GFR

K. Kunjuni Raja
University of Madras, India

Benson Mates
U. of Cal., Berkeley, USA

J.F. Staal
U. of Cal., Berkeley, USA

Pieter A. Verburg
Univ. Groningen, Netherlands

Editorial Assistant:
Arjen F. Florijn, University of Amsterdam, Netherlands

PUBLISHER: John Benjamins B.V.
Amsteldijk 44
AMSTERDAM, The Netherlands

SPONSORSHIP: The Foundation "Foundations of Language"

SUBSCRIPTION: Libraries and Institutions - Hfl. 90.- plus
Hfl. 12.- postage
Individuals - Hfl. 60.- (+Hfl. 5.-if payment is
by personal check)

PUBLICATION SCHEDULE: Three issues per year (c. 150 pages each).

First Issue: May 1977

Second Issue: August/September 1977

Third Issue: December 1977

ARTICLES, SQUIBS, AND DISCUSSIONS TO APPEAR IN:

STUDIES IN LANGUAGE, VOLUME 1

(three issues, all in 1977)

- Noel Burton-Roberts, "Generic Sentences and Analyticity"
- Robert M. W. Dixon, "Where have all the adjectives gone?"
- J. Durand, "Some aspects of reference and predication in the *Grammaire Générale et Raisonnée* and the Port Royal *Logique*"
- Bennison Gray, "Is there a science of parole?"
- Kwame Gyekye, "Akan language and the materialist thesis: a short essay in the relation between philosophy and language"
- John Haiman, "A study in polysemy"
- Jane Heal, "Ross and Lakoff on declarative sentences"
- Asa Kasher, "Bells in the gateway to language" [obituary article on the late Professor Y. Bar-Hillel]
- H. E. Hinch and Evelyn Pike, "Speaker-addressee versus third person within the Maung verb"
- Rodney Huddleston, "In defence of parasitic base structures"
- Y. Malkiel, "On hierarchizing the components of multiple causation"
- Joseph Margolis, "First- and second-language acquisition and the theory of thought and perception"
- Hans Moennig, "Meaning and speaker's intentions"
- Rita Nolan, "The semantics of 'possible': epistemic possibility"
- Vern S. Poythress, "Thirteen-box tagmemic theory as a method for displaying semi-independent language variables"
- Evelyn N. Ransom, "On the representation of modality"
- John Robert Ross, "Remnants"
- Geoffrey Sampson, "Reply to Graham"
- Benny Shanon, "Even, only, and almost hardly"
- Kenneth Shields, Jr., "Distinguishers and their implications for semantic theory"
- Georgette Silva and Sandra S. Thompson, "On the syntax and semantics of adjectives with 'it' subjects and infinitival complements in English"
- H. J. Verkuyl, "Thematic relations and the semantic representation of verbs expressing change"
- W. Zwanenburg, G. R. E. Ouwéneel and W. J. M. Levelt, "La frontière du mot en français"

ARTS & HUMANITIES CITATION INDEX (A&HCI)

Beginning in 1978 the Institute for Scientific Information (325 Chestnut St. Philadelphia PA 19106) will publish a new index, the ARTS & HUMANITIES CITATION INDEX, covering more than 1,000 key journals from literature, history, languages, religion, philosophy, drama/theatre, art, music, and other related fields.

Two softbound triannuals issued in June and October of each year will cover, respectively, the literature actually published from January through April and from May through August. A hardbound annual cumulation will appear the following May.

The ARTS & HUMANITIES CITATION INDEX will have three types of indexes: a citation index, an author index, a title-enriched key word index.

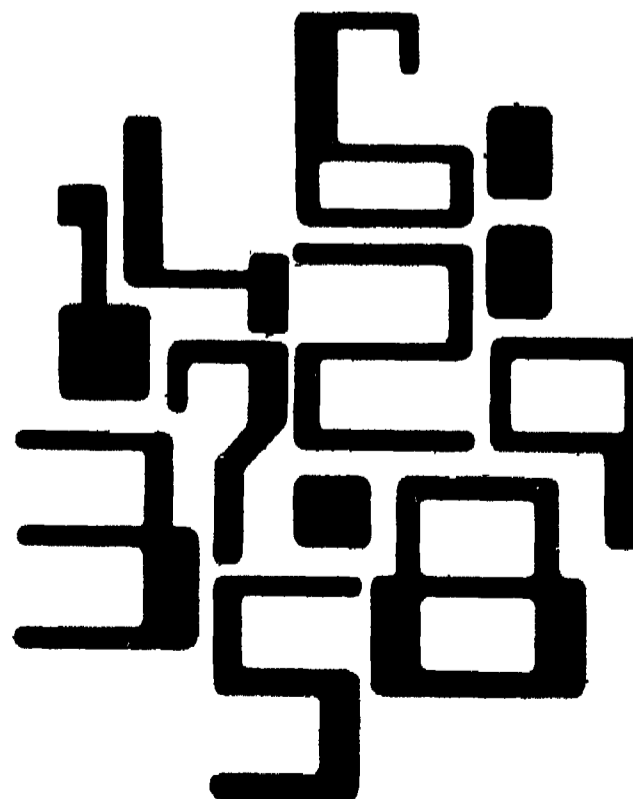
For more information: Ms. Susan Deutsch
Institute for Scientific Information
325 Chestnut Street
Philadelphia PA, 19106
(215) 923-3300

CURRENT BIBLIOGRAPHY

GENERAL 52
 PHONETICS-PHONOLOGY 53
 RECOGNITION 53
 LEXICOGRAPHY-LEXICOLOGY .. 54
 BILINGUAL 54
 STATISTICS 55
 DICTIONARY 59
 TEXT HANDLING 62
 GRAMMAR 63
 MORPHOLOGY 63
 PARSER 64
 GENERATOR 65
 CLASSES & CONSTRUCTIONS .. 65
 SEMANTICS-DISOURSE 68
 THEORY 70
 COMPREHENSION 71
 MEMORY 72
 TEXT GRAMMAR 73
 LINGUISTICS: METHODS:
 STATISTICAL 75
 ARTIFICIAL INTELLIGENCE .. 75
 MATHEMATICS 76
 SPEECH UNDERSTANDING 77
 DOCUMENTATION: INDEXING .. 77
 TRANSLATION 78
 SOCIAL-BEHAVIORAL SCIENCE:
 PSYCHOLOGY 83
 HUMANITIES 84
 ANALYSIS 86

In this issue and the next AJCL will be featuring abstracts of Russian language material covering the last four years. Our regular coverage will be back to normal with the next issue (the current 'slump' is the result of the pressures imposed on the bibliographer by dissertation writing).

AJCL thanks Martin and Iris Kay and Xerox Palo Alto Research Center for their help in preparing this bibliography.



GENERAL

Modelling the Contents and Meaning of Statements in a Logical Information System (O modelirovanii sodержaniia i smysla vyskazyvanii v informatsionno-logicheskoi sisteme)

V. D. Pistsov

Proceedings of the Scientific Seminar "Semiotics of the Means of Mass Communication" (Materialy nauchnogo seminara 'Semiotika sredstv massovoi kommunikatsii') Vol. 2, University of Moscow: 67-75; 1973.

Problems involved in developing a logical information system (LIS) as a model of human thought and speech activities are discussed. Requests written in a natural language are to be entered into the system, which is to produce answers in the same form. The LIS comprises three units: 1. receptor grammar--the input and request processing unit which dissects the request into individual statements; 2. memory; and 3. generative grammar--algorithm which forms the answer. Any statement is divided into content and meaning; meaning (-orientation) of a statement roughly corresponds to its division into theme and rheme, while content is tantamount to the statement after meaning has been excluded. It is supposed that the LIS will deal only with the content of statements, while the meaning of the answer will be determined by the request itself. 4 refs.

GENERAL

On the Informational Nature of the Applicative Generative Model (Ob informatsionnoi prirode aplikativnoi porozhdaiushchei modeli)

M. R. Melkumian

Proceedings of the Scientific Seminar "Semiotics of the Means of Mass Communications," (Materialy nauchnogo seminara 'Semiotika sredstv massovoi kommunikatsii') Vol. 2, University of Moscow: 25-36, 1973.

Shaumyan's applicative model is assessed from the standpoint of general semiotics. An interpretation of the applicative generation as "self-generating language system" is proposed. 13 refs.

A Device for Automatic Recognition of Phonemes (*Ustroivstvo dlia avtomaticheskogo raspoznavaniia fonem*)

I. a. Strel'nikov

Questions on the Analysis of Speech (Vopr. analiza rechi), 4 Tbilisi, "Metsniereba": 145-153, 1973.

A recognition device which isolates phonemes from the continuous speech flow. A phoneme is defined as a set of physical realizations grouped together by their common role in the verbal message in a given language. Each physical realization of a phoneme is described by a set of measurable features, which are the intensities in separate frequency bands of the acoustic spectrum. The device has 30 frequency bands encompassing the range from 60 cycles to 8.3 kHz. A recognition model is based on statistical theory of pattern recognition. A flow chart of the device is presented, describing the operations of the individual components, the input amplifier, the band filters, amplifiers of the filter channels, detectors, and the functional converter. 15 refs.

PHONETICS-PHONOLOGY: RECOGNITION

A Method for Modification of the Overall Intensity of Sound Intervals for Machine Recognition (*Ob odnom sposobe izmeneniia abshchei intensivnosti zvukovkh otrezkov vtseliakh ikh mashinnogo raspoznavaniia*)

I. a. Strel'nikov

Questions on the analysis of Speech (Vopr. Analiza Rechi 4 Tbilisi "Metsniereba": 154-159, 1973.

The operation of the recognition automaton includes a procedure which prepares the description of the pattern in such a way as to exclude the influence of isomorphic transformations, that is the changes of the input image. A method is proposed which makes the recognition procedures independent of the isomorphic transformations. The concept is based on a periodical modulation in the range of existence of the pattern or on a periodical modulation of all the parameters distinguished. Applications of this procedure are discussed with special reference to the isomorphic transformation of the type of "general intensity modification" of speech flow for a probabilistic recognition model. 5 refs.

Lexical Combinability as an Object of Bilingual Lexicography (*Leksicheskaia sochetanost' kak ob'ekt dvuiazychnoi leksikografii*)

N. I. Sukalenko

Mathematical Linguistics (Mat. lingvistika), t. 1. Kiev University: 110-117. 1973

Applications in bilingual dictionaries of lexical functions (LF) proposed by A. K. Zholkovskii and I. A. Mel'chuk are discussed. The author suggests certain refinements of LF and the addition of new, more specific LF to the existing set. For example, the LF with the meaning, "to show to advantage" for describing such expressions as *to shine with knowledge, to set off whiteness, to strike the superiority, to single out for ability*; the LF with the meaning "diverge from the ordinary course of events" for such phrases as *to lose (the way, course), to get off (the subject), to diverge (from the conversation), to jump (the tracks), to turn off (one's route), to drift away (from the old friends)*. Apart from standard LF, the entries of bilingual dictionaries must contain individual LF *service--re-engaged, active*, etc. and information about semantic combinabilities: *fat* only in relation to people and animals, *stout*--only about people, *buxom*--only about women.

LEXICOGRAPHY-LEXICOLOGY: BILINGUAL

German-Russian Automatic Dictionary of Common Words (*Nemetsko-russkii avtomaticheskii slovar' obshchepotrebitel'noi leksiki*)

N. E. Okulich

Linguostatistics and the Automated Analysis of Texts (Lingvostatistika i avtomat. analiz tekstov), Minsk, 354-382, 1973.

An alphabetic listing of 1681 words is given, considered as the common vocabulary for scientific and technical German prose. The words were selected on the basis of comparing frequency word lists for six subsets of technical style: words which had the same meaning in all the six subsets were included in the list. Some specific selection criteria are discussed (e.g. only the words which were not rare in the frequency dictionaries of the usual type were counted as common words). The list agrees well with similar lists for English, earlier compiled by the Speech Statistics Group. The comparison was based on German-English and English-German dictionaries. 13 refs.

The thesaurus method of Automatic Recognition of the Semantic Pattern of Scientific and Technical Text (*Tezaurusnyi metod avtomaticheskogo raspoznavanija smyslovogo obraza nauchno-tekhnicheskikh tekstov*)

A. N. Popesky, and M. S. Khazhinskaia

Linguostatistics and the Automated Analysis of Texts (Lingvostatistika i avtomat. analiz tekstov), Minsk: 296-324, 1973.

With special reference to the drying of varnish and paint coating, a bilingual French-Russian thesaurus and an auxiliary thesaurus of relators have been constructed. The special thesaurus includes a graphic representation of the connected descriptors and the so-called "thesaurus forms", in which the French terms are accompanied by Russian equivalents. The auxiliary thesaurus is a list of 57 stencil relators and indicator relators, which differ by the degree, of abstraction and are joined together into 11 topics expressing the semantico-syntactic relationships between the descriptors, such as equivalence, order, dependence, etc. The input and organization of the thesaurus in the Minsk-22 computer memory are described as is the operation of the recognition algorithm working on the basis of the thesaurus.

LEXICOGRAPHY-LEXICOLOGY: STATISTICS

Word Lists of the Electrotechnical and Radio-engineering Subsets of Language: Compilation Procedures and Efficiency Testing (*Methodika sostavlenija slovnika pod"iazyka elektro- i radiotekhniki i proverka ego effektivnosti*)

K. F. Pruzhina

Mathematical Linguistics (Mat. lingvistika), T. 1. Kiev University: 74-79, 1973.

A word list comprising 3268 words and 447 set nonterm phrases has been compiled on the basis of French texts in electrical and radio engineering. Data on the rate of appearance of new words in compiling the word list and on distribution of the words by the groups are given. The groups are as follows: (1) words of a common stem with other words already included in the list; (2) international words having the same meaning as the corresponding Russian words; and (3) the rest of the words. Statistical calculations and testing of the word list by texts in the special field under consideration have shown the word list to cover practically all French electrical and radio engineering texts. 9 refs.

Using Statistical Methods in Lexicological Research (with Special Reference to Rumanian) (Ob ispol'zovanii statisticheskikh methodov v leksikologicheskom issledovanii. (Na materiale leksikostatisticheskikh issledovanii rumynskogo iazyka))

S. B. Semchinskii

Mathematical Linguistics (Mat. lingvistika), t. 1., Kiev University: 86-96, 1973.

The findings of lexical statistical studies of Rumanian etymology are presented as obtained by Rumanian authors (D. Copceag, c. Dimitriu, C. Tudose and C. Maneca) in recent years; in particular, data are given on the proportions of Roman and non-Roman words in Rumanian manuscripts of different epochs and on the proportions of motivated/nonmotivated words comparatively for Rumanian and Russian. Shortcomings of the studies are pointed out, specifically the limited input material, some mistakes in descriptions of the etymology and morphemic composition of words (for instance, of the 280 nonmotivated words which D. Copcaeg treats as purely Roman, more than 60 are words of a Slavonic origin; in discussing Russian words, he treats as motivated such simplified words as *fiagushka*, *schet*, *vereteno lestnitsa*)

LEXICOGRAPHY-LEXICOLOGY: STATISTICS

Vocabulary of One Day: Derivational Frequency Guide to Ten Newspapers (Slovar' odnogo dnia (chactotnii slovoobrazovatel'nyi spravochnik po desiati gazetam))

B. N. Iakolev

*Mathematical Linguistics Chair
Leningrad University*

Proceedings of the Scientific Seminar "Semiotics of the Means of Mass Communication," (Materialy nauchnogo seminara- 'Semiotika sredstv massovoi kommunikatsii') Vol. 2, University of Moscow: 194-198, 1973.

Within the framework of Russian frequency dictionary project, a frequency derivational dictionary to ten Soviet newspapers in Russian, dated January 5, 1968, is being compiled.

**An Essay on Compiling a Frequency Dictionary of Modern English
Combinabilities (*Opyt sostavleniia chastotnogo slovaria sochetaemosti
sovremennogo angliiskogo iazyka*)**

S. S. Khidekel', N. O. Volkova, R. S. Genzburg, V. I. Perebeinos, D. A. Sankin, and D. A. Sankin

Problems of Lexicology (Problemy Leksikologii), Minsk, Belorussian University: 180-191, 1973.

The current interest is stressed of studying lexical combinabilities both in a general theoretical aspect and for a number of applied linguistic studies, primarily for a scientific selection of the minimum vocabulary in teaching a foreign language. Combinability can be studied by linguistic and statistical methods. The latter imply compiling a frequency word list of phrases. Difficulties involved in compiling such word lists are discussed. The main one is the necessity of investigating a very large text file. A plan for compiling a dictionary of English phrases under development at the Institute of Foreign Languages, in Moscow, jointly with the Institute of Linguistics of the Ukrainian Academy of Sciences, is reported. The dictionary is planned to give combinability data for 1,000 most frequent words of modern English.

LEXICOGRAPHY-LEXICOLOGY: STATISTICS

First-letter Word Statistics in Russian Printed Text (*Statistika slov po nachal'nym bukvam v russkom pechatnom tekste*)

Iu. A. Saf'ian

Proceedings of the Scientific Seminar "Semiotics of the Means of Mass Communication (Materialy nauchnogo seminara 'Semiotika sredstv massovoi Kommunikatsii'), Vol. 2, University of Moscow: 199-209, 1973.

Data on quantities and frequencies of different words beginning with a certain letter in three styles of Russian--Technical (T), natural science (NS), and fiction (F)--are given in four tables and a diagram as obtained during a study of the respective frequency lists. The relative disparities in the number and frequency of different words beginning with some one letter between the dictionaries T and NS, NS and F are given (Table 4).

**Distributional Statistical analysis of Semantics of a Group of Words
(Distributivno-statisticheskii analiz semantiki odnoi gruppy slov)**

L. T. Toroshechnikova

Problems of Lexicology (Problemy leksikologii), Minsk, Belorussian University: 170-174, 1973.

The findings of a statistical analysis of phrases including lexemes denoting persons show that 21 nouns have the highest frequency characteristics: *baby, boy, child, chap, friend*, etc. From texts by modern English writers (totalling 25 million running words) some 150,000 phrases with the nouns have been extracted. All the phrases were divided by the subject principle into 103 classes (the classes were based on Roget's thesaurus, Kunin's phraseological dictionary and author's intuition). Statistical distributions of the occurrences of analysed words in the subject classes were compared and the indices of semantic proximity were calculated for pairs of words (the highest proximity was found to relate the following words: *boy--girl; child--boy; man--woman; mate--kid; lad--chap*, and the lowest proximity, for the following words: *human being--friend; baby--people; citizen--friend*) 4 refs.

LEXICOGRAPHY-LEXICOLOGY: STATISTICS

The Frequency Dictionary of the Metallurgy Subset of English (Chastotnyi slovar' angliiskogo pod''iazyka metallurgii)

N. V. Gurova

Theory of Language and Engineering Linguistics (Teoriia iazyka i inzh. lingvistika), Leningrad; 94-102, 1973.

A word frequency list of 1,024 most frequent lexemes is given as occurring in the sample of text on metal rolling from British and US periodicals covering 1965-1971. The sample contained 188,000 running words, amounting to 5,300 different lexemes (10,500 different text words). The lexemes in the list had a frequency not under 22; the rest of the data are given in the frequency distribution of lexemes (there were 1,500 lexemes with the frequency 1 in the sample; 683 with the frequency 2, etc.).

Byte Program for Frequency-Alphabetic Dictionary (*Programma postroeniia chastotno-alfavitnogo slovaria na mashine sbaitovoi strukturoi*)**T. P. Karnilovich***Linguostatistics and the Automated Analysis of Texts (Lingvostatistika i avtomat. analiz tekstov), Minsk: 450-459, 1973.*

The features of a byte-memory computer are described. The basic advantage of the system is the fact that for access to information on the disk (drum) it is not necessary to scan all the preceding information as with the magnetic tape. This enables an essential increase of the volume of the input text in frequency dictionary compilation (a text of up to 500,000 running words can be processed simultaneously; at present text is entered by portions of 1,000--5,000 running words). Correspondingly, operation time is cut down. A schematic algorithm for alphabetic-frequency dictionary compilation is given, as is algorithm and program of one of the components of the schematic algorithm. 4 refs.

LEXICOGRAPHY LEXICOLOGY: DICTIONARY

Computer-based Machine and Human-readable Multilanguage Dictionaries and Multi-Aspect Reference Guides, with Special Reference to the Code List of Fish, Benthos, and Plankton (*Razrabotka sispol'zovaniem E.V.M. mashino- i cheloveko-vostrinimaemykh mnogoiazychnykh slovarei i mnogoaspectnykh spravochnikov na primere reestrovogo kodifikatora ryb, bentosa i planktona*)**V. A. Kontor***Proceedings of the Scientific Seminar "Seminar "Semiotics of the Means of Mass Communication, (Materialy nauchnogo seminaru "Semiotika sredstv Kommunikatsii") Vol. 2, University of Moscow: 210-216, 1973.*

The author discusses the difficulties to be faced in compiling dictionaries and reference aids that can be used both by humans and by the computer. A code list of fish, benthos and plankton encompassing 50 million symbols has been compiled by means of a computer. Each line of the list carries an ordered sequence: the code of the species; the Latin name of the species; the name of the author who described this species, etc. The software for correcting the file in the computer memory comprises eight routines: 1. Control of the block boundaries. This routine estimates the general length and format of the block of input data first submitted. 2. Position control. Verifies the internal structure of the block. 3. Grammar control. Formats the machine memory and prints out the first version of the reference list intended for human users. Galleys are produced and submitted to a specialist for editing. The remaining five routines eliminate unnecessary duplication of data arising in preparing and inputting the file.

Hierarchical Relations of Nominative Units of the Mass Media Language (*Ierarkhicheskie otnoshenija nominativnykh edinits iazyka S.M.I.*)

G. A. Smirnova

Proceedings of the Scientific Seminar "Semiotics of the Means of Mass Communication (Materialy nauchnogo seminaru "Semiotika sredstv massovoi kommunikatsii"), Vol. 2, University of Moscow: 217-230, 1973.

A model of a natural language dictionary is proposed. For describing a dictionary, a hierarchical calculus of specially constructed objects ("nomemes") is built as well as a system of rules for transition from these objects to the real vocabulary units. Each word or phrase is a production in that calculus. The unit objects are predicates, actants and operators. All these are specified in a list, for example, X_i is the trivalent predicate having instrumental, object and subject valencies; Y_i is instrumental actant, D^1 and D^2 are depredicator-operators. The input formulas--predicate structures--are built out of predicates and actants. For example, X_i is a trivalent predicate with the instrumental valence realized. By applying operators to the initial formulas, other formulas are built. For instance, if X is a formula, D^1X and D^2X are formulas. (The rules for transition from nomemes to the units of the natural language are not given.--*Abstractor's note.*)

LEXICOGRAPHY-LEXICOLOGY DICTIONARY

The Information Bank of Dictionaries (*Informatsionnyi bank slovarei*)

L. N. Zaserina, and P. V. Sil'vestrov

In Proceedings of the Scientific Seminar "Semiotics of the Means of Mass Communication" (Materialy nauchnogo seminaru "Semiotika sredstv massovoi kommunikatsii"), Vol. 2, University of Moscow: 217-230, 1973.

Dictionaries are proliferating in number and type and bibliography on them is lagging. A semi-automatic dictionary seems to be a way to begin solving the problems this situation presents by initiating a gradual transition to a comprehensive system for the registration of lexico-grammatical data - an information bank of dictionaries (IBD). The IBD is to be supported by a network of lexicographic centers which are to channel their results to the head center where the computer will transform the information by standard routines. Two additional systems are necessary: a) information retrieval catalogue and b) an information retrieval language compatible with it to provide for automatic question answering. The IBD is to consist of a few more or less autonomous subsystems so that it can be developed by individual components; one of the first tasks to be fulfilled may be the integration of word lists of the general dictionaries of Russian and cumulation of the lexical features ascribed to each word. A possible IBD is described.

LEXICOGRAPHY-LEXICOLOGY: DICTIONARY

Essay on Computer Implementation of a Distributional Method for Determination of Lexical Meanings (*Opyt mashinnoi realizatsii distributivnoi methodiki opredeleniia leksicheskikh znachenii*)**I. N. Marchuk***Statistics of Speech and Automated Analysis of Texts (Statistika rechi i avtomaticheskii analiz teksta), Science, Leningrad, 181-230, 1972(1973).*

Every polysemous word in the dictionary has a set of diagnostic forms associated with it which discriminate its meanings and a set of translation equivalents uniquely associated with these forms. Given algorithms for recognition of diagnostic forms in the text, this would suffice for homonymy resolution. Such a dictionary was built for English newspaper texts (830,833 text words, 30% having multiple meanings) by comparing and gradually improving a concordance dictionary based on a bilingual text and giving the distribution, frequency, and all meanings of each word. The meaning resolution algorithms take into account morphological features, lexical-grammatical class and syntactic functions of the word in the sentence (determined without semantic criteria). Sets of words or features which were detected in the context and enable a unique determination of one of the meanings are determinants. There are five types of simple determinants; with compound determinants built from simple ones. For a majority of verbs one or two determinants are used. A machine translation experiment using this system on 30,000 running words of text has been carried out, yielding correct translation of multiple-meaning words in 80% of cases. Samples are given.

LEXICOGRAPHY-LEXICOLOGY: DICTIONARY

Choosing the Type of the Output Automatic Dictionary (*Vybo tipa vykhodnogo avtomaticheskogo slovaria*)**V. A. Veltel'***Linguostatistics and the Automated Analysis of Texts (Lingvostatistika i avtomat. analiz tekstov), Minsk, 383-401, 1973.*

The characteristics of two types of automatic dictionary are compared--word form dictionary (WFD) and stem dictionary (SD). WFD is shown to contain far more redundant information than SD; a formula expressing this difference in quantitative terms is given. These two types of dictionary are further compared by criteria of I. Mel'chuk--exhaustiveness, adequacy, economy, simplicity, and compactness. A set of indicators to evaluate each of the dictionaries is introduced. With special reference to dictionaries of several finite text samples, the relationship between WFD and SD according to Mel'chuk's criteria is shown: the stem dictionaries are substantially better as regards exhaustiveness, economy and compactness, but they are inferior to word form dictionaries in adequacy and simplicity. The choice of the dictionary type is determined, over and above these relationships, by such factors as computer design, input and output languages, etc.

A Method of Automatic Establishment of Lexical Uniformity of Text (*Ob odnom sposobe avtomaticheskogo vyivleniia leksicheskoi odnorodnosti teksta*)

K. F. Luk'ianenkov

Linguostatistics and the Automated Analysis of Texts (Lingvostatistika i avtomat. analiz tekstov), Minsk: 325-338, 1973.

The lexical uniformity of a sample is considered to be no less important than its size, representativeness, etc. A method for determination of the lexical uniformity of text with the aid of the computer is proposed (general flowchart of the algorithm is given). The method is based on statistical evaluation of the difference of the relative frequencies of word forms occurring in two (or more) portions of text to be compared as well as on the basis of the evaluation of the general amount of coincident word forms. The application of the method is illustrated by comparison of two portions of a hypothetical text. The word list of one portion covers up to 96% of the word occurrences in the other portion; only 70% of the coincident words have statistically insignificant differences in the relative frequencies. This result is regarded as sufficient for concluding as to the lexical uniformity of text represented by the two portions being compared. 13 refs.

LEXICOGRAPHY-LEXICOLOGY: TEXT HANDLING

On Some Procedures for Obtaining Reduced Word Codes (*O nekotorykh metodov polucheniiia svernutykh kodov slov*)

G. Z. Mikhtin, R. G. Plotrovskii, and V. A. Frumkin

Linguostatistics and the Automated Analysis of Texts (Lingvostatistika i Avtomat. analiz tekstov), Minsk: 286-295, 1973.

The importance of reducing the volume of input information by efficacious coding is stressed. The general techniques of word form code reduction are discussed. The techniques are shown to be efficient for reducing the length of input text but they fail to provide for unambiguous identification of the word form. It is proved that in principle such a solution of the reduction problem exists which ensures unambiguous identification. This solution is supposed to be based on statistical-distribution features of text. 4 refs.

Segment Analysis of Verbal Word Forms in Russian (*Segmentnyi analiz glagol'noi slovoformy v russkom iazyke*)

L. A. Chizhova

Moscow University Herald. Philology. No. 4: 3-11, 1973. (Vestnik Maskovskogo universiteta. Filologiya.)

Segment analysis of Russian verbal word forms can proceed in three stages: 1) isolation (in the word form) of segments, 2) determination of the syntagmatic position of the segment depending on the meaning it expresses and establishment of the type of meanings expressed by the segment, (in this way morphs are singled out in the word form), 3) merger of individual positions in one order if their meanings coincide, but independent of the expression of these meanings. An order is defined as the sum of positions of the morphs that are combined within one morpheme. Nine orders are introduced for personal verbs; they are linearized in strict sequence. Compulsory/optional presence, in a word form, of morphs of this or that order and combinabilities of individual orders are discussed. Conclusions regarding the dependence of the meaning of a morph on its position in the word form are made. Examples of segment analysis of verbal words forms are given.

GRAMMAR: MORPHOLOGY

Construction of a Morphological Analysis Algorithm for Russian on the Basis of a Finite Automaton Model (*Postroenie algoritma morfologicheskogo analiza dlia russkogo iazyka na osnove konechnoavtomatnoi modeli*)

G. S. Ul'ianova

Linguistic Problems of Functional Modeling of Speech (Lingvisticheskie problemy funktsional'nogo modelirovaniia prechevoi deiatel'nosti), Leningrad University: 145-174, 1973.

An algorithm of morphological analysis of Russian is described that has been developed at the Mathematical Linguistics Group of the Computing Centre of Leningrad University. The input data for the algorithm are the Russian entry word form and a tabular dictionary describing the Russian morphology in terms of a finite automaton model. The tabular dictionary falls into a dictionary of stems and the table of affixes. The algorithm scans the alternative paths for generation of the word form. The structure of the tabular dictionary is described (each line contains a string generated) as are the preceding and subsequent states of the automaton. A list of morphological features for different parts of speech is given. The operation of the algorithm is illustrated by the word form *vyt solniaemyi* fulfilled. A computer experiment is described that has been carried out on a text of 80 sentences--982 word forms. 518 words were processed correctly, 442 not processed at all as the words were missing in the dictionary and 22 words (four of these occurred twice) received 1 or 2 excessive analyses. A fragment of the dictionary of stems and complete dictionary of affixes are given (347 inflections of *S*, *A* and *V*).

On Syntactic Analysis of Context-Free Languages (*k voprosu o sintaksicheskoi analize kontekst-svobodnykh-iazykov*)

A. Sh. Nepomniashchaia

In A Few Questions on Theoretical Cybernetics and Algorithms of Programming (O nekotorykh voprosakh teoreticheskikh kibernetik i algoritmakh programmur), Novosibirsk: 134-144, 1971.

An algorithm for syntactic analysis of a subclass of context-free languages is proposed which is represented by means of a pushdown nondeterministic automaton. The task of syntactic analysis of a context-free language is to define, for each string of the language, its structure, that is to define the sequence of context-free grammar rules producing that string. A subclass of context-free grammars is discussed whose rules meet a number of restrictions. First, the rules of these grammars are defined in the Rosenkranz normal form: a, b, c are the terminal symbols, x is a nonterminal symbol, w is a string of auxiliary symbols; second, a number of restraints of the following type are formulated: if the grammar has a rule of a certain form, the presence of a rule of a certain other specified form is prohibited. A simple procedure is established for a given class of context-free grammars, which for a given class of context-free grammars, which for each input string x_0 from $I(x_0)$ -- the set of alternative analysis of the string x_0 , as produced by the algorithm handling the representation of the string x_0 in the nondeterministic pushdown automation) chooses only alternatives which are correct analyses of the string x_0 . 9 refs.

GRAMMAR: PARSER

Description of an Experiment in "Simplified" Syntactic Analysis of Scientific Russian (*Opisanie eksperimenta po "uproshchennomu" sintaksicheskoi analizu russkikh nauchno-tekhnicheskikh tekstov*)

M. S. Pershikova, and T. D. Lagacheva

Linguistic Problems of Functional Modeling of Speech (Lingvisticheskie problemy funktsional'nogo modelirovaniia rechevoi deitel'nosti), Leningrad University, 102-110, 1973.

The procedure of automatic syntactic analysis is to be broken down into two phases: preliminary and extended. The preliminary phase of the syntactic algorithm is described as carried out on Russian material. This phase makes use of information about words and a "minigrammar"--an abridged dependency grammar. Information on words (a word is defined as a set of homonyms) comprises the following three groups of features: I. Word classes (part of speech). II. Word class restrictors. III. Syntactic features. The "minigrammar" is based on 8 conventions as to the directions of arrows in a dependency tree. The set of conventions disregards difficult cases (this is done expressly). The grammar comprises 202 rules. The preliminary algorithm produces the following information: 1. it establishes whether the sentence has at least one acceptable analysis; 2. if so, which types of government occur at least in one of the acceptable analyses and which homonyms are apices at least in one of acceptable analyses; 3. which of the homonyms are chosen in at least one acceptable analysis. A computer testing of the algorithm on 80 Russian sentences from an abstract journal in welding has been carried out. Of these, 70 sentences were processed (the analysis of one sentence takes some 30 sec). The mistakes that were detected during the experiment have not been corrected in the grammar because of some technical reasons.

Programming a Machine Translation Problem: A Case Study
 (Programmirovaniia zadach mashinnogo pervoda)

D. Ia. Levin, and G. I. Nekrasov

In A Few Questions on Theoretical Cybernetics and Algorithms of Programming (O nekotorykh voprosakh teoreticheskikh kibernetik i algoritmakh programmur), Novosibirsk: 145-158, 1971.

Implementation of an algorithm of morphological synthesis of words and linearization of words in a sentence on the basis of that sentence's syntactic structure is considered. An exact statement of the problem is given, and the algorithm is described in a verbal form. The program implementing this algorithm is presented in full; it is written in an Algol-type language, specially developed for machine translation problems. The program includes detailed comments specifying the coding and implementation. 4 refs.

GRAMMAR: CLASSES & CONSTRUCTIONS

Transformations in Symmetric Structures: Coordination and Ellipsis
 (Transformatsiia vsimmetrichnykh konstruktsiiakh konstruktsiiakh: sochinenie i ellipsis)

T. D. Korel'skaia, and E. V. Paducheva

Scientific Technical Information, Collection, All-Union Institute of Scientific and Technical Information (Nauchno-tekhnicheskais informatsiia. Sbornik. Vse-soiuznyi institut nauchnoi i tekhnicheskoi informatsii), No. 9: 29-38, 40, 1973.

A formal description is offered of two syntactic transformations--conjunction reduction and symmetric ellipsis. The applicability of the two transformations is shown to be determined by the analysis of the sentence (or its fragment) into immediate constituents. The concept of functional sentence phrase structure is defined as based on the dependency tree, word order, anaphoric identity relationship and membership of a symmetric structure. The importance of the tone pattern for statement of the applicability conditions of both the transformations has been established. The conjunctive and elliptic reductions are shown to be independent of the grammatical nature of the constituents. The conjunctive reduction is extended to the case of symmetric but not necessarily coordinated parts of the sentence. Similarities and distinctions between conjunction and symmetric ellipsis are characterized. Problems of representation of the structures of coordinated and elliptic sentences are discussed. 24 refs.

Ways of Production of Constructions with Assimilating Pronouns in the Inner Text Structure (Sposoby razvertyvaniia konstruktsiy s upodobliaiushchimi mestoimeniiami vo vnutrenney strukture teksta)

M. I. Otkupshchikova

Linguistic Problems of Functional Modeling of 'Speech (Lingvisticheskie problemy funktsional'nogo modelirovaniia rechevoi deitel'nosti), Leningrad University: 132-140, 1973.

The elements of the inner text structure are objects and functors; among the latter, predicates are distinguished, serving to link objects, and connectors, serving to link situations (a situation is a predicate with places filled in by objects). For automatic semantic analysis the ways of expression of similarity functor in Russian are investigated. Two types of similarity functor are distinguished depending on the structure of its places: three-place similarity predicate ("A is similar to B in respect of c") and similarity connector ("A is similar to B", where A and B are situations). One of the means of expression of the similarity functor is special pronouns, which are referred to as assimilating pronouns (*the same, identical to, as many, in the same place*, etc.). The author proposes to describe the meaning of assimilating pronouns by means of a partially hierarchized set of binary features.

GRAMMAR: CLASSES & CONSTRUCTIONS

An Essay on Compound Sentence Segmentation on the Basis of Boundary Punctuation (Opyt segmentatsii slozhnykh predlozheniy na osnove organichitel'nykh znakov prepiniia)

A. G. Davtian

Linguistic Problems of Functional Modeling of Speech (Lingvisticheskie problemy funktsional'nogo modelirovaniia rechevoi deitel'nosti). Leningrad University: 26-33, 1973.

The functions of semicolon and colon in English radio-location texts (9,000 sentences) are examined. Semicolon either separates sentences or connects coordinate noun phrases and infinitive phrases. The block of preliminary sentence segmentation must take account of the former case and disregard the latter one; rules for sentence recognition are used. As to colon, four main types of its usage in text have been identified: 1. after margin heading to separate it from the following text; 2. preceding an enumeration of coordinate parts of sentence preceded by a generalizer word; 3. between parts of a compound sentence not connected by a conjunction (has the same function as a semicolon); and 4. after words introducing direct speech.

A Type of Substitute Word in Connected Text (Ob odnom tipe slov-zameštiteley vsviaznom tekste)

L. N. Smirnova

Linguistic Problems of Functional Modeling of Speech (Lingvisticheskie problemy funktsional'nogo modelirovaniia rechevoi deitel'nosti), Leningrad University: 63-75, 1973.(635)

For automatic analysis of Russian one type of substitutes in anaphoric relation is analyzed, where the substitute word is a noun (sometimes with an identifying modifier). For the identifier, indicative and possessive pronouns, numerals, adjectives and participles are used. Depending on the semantic relationship of the concepts expressed by the main word (apex) of the antecedent, on the one hand, and the noun in the substitute on the other, four types of substitutes are distinguished. 1. The noun in the substitute (S') coincides with the apex (A) of the antecedent. 2. S' is a syntactic derivative of A expressed by a verb, participle or adjective. 3. S' is a classifier (i.e., a generic term) with respect to A . This type covers the contextual synonyms that can be identified using lexical functions. Estimating modifiers may occur in all the above types in S' . On the basis of this classification, conditions are formulated for establishing anaphoric relations in case the apex of the substitute is expressed by S .

GRAMMAR: CLASSES & CONSTRUCTIONS

On Translation from a Logical Information Language into Russian: Transformational "Insertion of Classifiers" (K' problem perevoda s informatsionno-logicheskogo iazyka na russkii: preobrazovanie "vvedenie klassifikatororov")

G. E. Kreidlin

Scientific-Technical Information, Collection, All-Union Institute of Scientific and Technical Information (Nauchno-tehniskaia Informatsiia. Sbornik. Vse-soiuznyi Institut Nauchnoi i Tekhnicheskoi Informatsii) Series 2, No. 12: 14-19, 39, 1973.

The author discusses the investigation of some correlative words in Russian--*fakt* 'fact', *abstoiatel'stvo* 'circumstance', *utverzhdenie* 'statement'. The meanings of these language units are described through the rules by which they are introduced into the sentence structure; by which they are interpreted through structures which belong to the logical information language and are treated as elementary. The conditions are described under which the transformation of "insertion of classifiers (correlative words)" can be performed describing the changes that it involves in the sentence structure. For the word *utverzhdenie* additional synonymic transformations are described (with a special reference to mathematical text) illustrating the possibility of varying the syntactic structures containing that word. 11 refs.

A Method for Description of Semantic Units (Ob odnom metode opisaniia semanticheskikh edinits (mnozhestvennye ob"ekty smyslovoi struktury))

B. M. Leikina

Scientific-Technical Information, Collection, All-Union Institute of Scientific and Technical Information, Series 2, No. 8: 3-12, 46, 1973 ("Nauchno-tekhnicheskaiia informatsiia. Sbornik. Vse-soiuznyi institut nauchnoi i tekhnicheskoi informatsiia," 1973, seriia 2)

Semantics natural language is described. The categories of plurality and its expressions in the language are analyzed. 21 refs.

SEMANTICS-DISOURSE

Semantics of Information Languages (Semantika informatsionnykh iazykov)

G. M. Il'in

In Linguistic Problems of Functional Modeling of Speech, Leningrad University: 33-37, 1973. ("Lingvisticheskie problemy funktsional'nogo modelirovaniia rechevoi deiatel'nosti". 1.L. leningradskii universitet)

An information language (IL) must meet the following requirements, 1. The contents of a document (or a fact, in case of a factographic information retrieval system), as written in the IL units, must be represented with sufficient completeness. This requirement poses the problem of choice of descriptors and grammar. 2. IL must provide for formal definition of the transformation rules of language utterances (the rules for synonymic transformations and rules of logical sequence). The latter requirement determines the former one. IL must reflect only those semantic differences as can be taken into account by transformation rules. For example, the description of the meaning of the word *tol'ko* (only) is.

$$P(a, \text{tol'ko } b) - P(a,b), P(a; \text{tol'ko } b), - \text{INCL } (c,b) - -P(a,c),$$

where INCL is the "inclusion" predicate and - is the negation sign.

Three Levels of the Semantic Structure of Sentences (*Tri urovnia semanticheskoi strukmuri predlozheniia*)

G. G. Pecheptsov

In Mathematical Linguistics (Mat. Lingvistika), Kiev University: 66-73, 1973.

The author proposes to analyze the semantics of a sentence from three viewpoints: (1) as the sum total of the meanings of smaller constituents--words and phrases; (2) as a constituent of a higher level (sentences are subdivided into those carrying information and those devoid of it); and (3) from the standpoint of the "superstructure". The superstructure is associated with the meaning of the sentence as a whole and comprises three components: (a) "central idea" (the theme and the rheme); (b) context (illustrated by pairs of sentences of the question-answer type); and (c) presupposition. 27 refs.

SEMANTICS-DISOURSE

Applicative Grammar as a Universal Semiotic System Modelling Natural Languages (*Applikativnaia grammatika kak universal'naia semanticheskaiia sistema, Modeliruiushchiia estestvennye iazyki*)

S. K. Shaumian

Proceedings of the Scientific Seminar "Semiotics of the Means of Mass Communication" (Materialy nauchnogo seminara 'Semiotika sredsiv massovoi kommunikatsii') Vol. 2, University of Moscow: 5-10, 1973.)

An applicative genotype grammar (AGG) is a universal system which serves as the basis of all NLs. By imposing restraints on the AGG, one can obtain grammars which generate phenotypic languages (particular NLs). The explanatory power of the AGG lies in its universality while its predictive power lies in its ability to predict as yet unobserved properties of NL's. The theoretical study of NL's implies: 1) reconstruction of a genotype language which exists objectively but is not liable to immediate observation, 2) construction of the grammar of this language, 3) construction of the derivative genotype grammars, 4) the study of the transformations which give rise to phenotypic languages, 5) the construction of a typology of NLs on the basis of these transformations, 6) investigation of the laws governing these semiotic systems, 7) explanation of the transformations of the genotype language from the viewpoint of these laws and the prediction of the possible types of semiotic systems.

Deep Structures and Semantics (*Glubinnye struktury i semantika*)

V. V. Bogdanov

Theory of Language and Engineering Linguistics (Teoriia iazyka i inzh. lingvistika), Leningrad: 134-145, 1973.

The relationship of syntax and semantics in some modern linguistic theories is discussed (transformational generative grammar, and interpretative semantics, generative semantics; Ch. Fillmore's case grammar, and the "meaning \leftrightarrow text" model, etc.). Different interpretations of "deep structure" and "surface structure" are considered. The author suggests assigning the term "deep" to semantic structure and "surface" to syntactic structure of the sentence.

SEMANTICS DISCOURSE: THEORY

A Universal Language for Formal Description of Physical Laws (*Universal'nyi iazyk dlia formal'nogo opisaniia fizicheskikh zakonov*)

P. G. Kuznetsova

Proceedings of the Scientific Seminar "Semiotics of the Means of Mass Communication (Materialy nauchnogo seminara 'Semiotika sredstv massovoi kommunikatsii'), Vol. 2, University of Moscow: 141-161, 1973.

To map an intuitive theory into a mathematical one, correct definition of terms is required. A term in physics should be treated as correct if and only if for the quantity it denotes there exists a measurement procedure which sets into correspondence with that term a reading of an instrumental scale at each time point. This makes it possible to identify every term with a coordinate axis of some hypothetical coordinate system. The axioms of the theory in that case would be statements about instrument readings; these should be written down in an invariant form, that is in a notation independent of the coordinate system adopted (tensor analysis). The notions of the movement of solids, distance between two points, length, area and volume are discussed, demonstrating that they can be represented in the language proposed. A definition of the physical quantity and physical law is given, specifying the sequence of operations in mapping an intuitive theory into a mathematical one. To ensure a uniform operational interpretation of all physical laws through measurement procedure the author proposes to eliminate mass from the set of dimensional values which is bound to lead to a drastic breakdown of the entire system of dimensionalities.

SEMANTICS-DISOURSE: COMPREHENSION

Identification and Linking of Variables During Semantic Analysis of Mathematical Texts (*Otozhdestvenie i svlazyvanie peremennykh pri semanticheskoi analize matematicheskikh tekstov*)

V. D. Butonov

Linguistic Problems of Functional Modeling of Speech (Lingvisticheskie problemy funktsional'nogo modelirovaniia rechevoi deitel'nosti), Leningrad University: 11-26, 1973.

To fix both the intermediate and end result of semantic analysis a generative semantic model (SM) is used; the resultant semantic representation (SR) of the sentence being obtained recursively through the SR of its constituents. The language of SM is a syntactical extension of the language of narrow predicate calculus. The input to SM is a dependency tree of the sentence in which anaphoric relationships have been restored. The nodes of the tree are labeled with symbols of dictionary SM representation (which includes syntactic valencies) of the corresponding words and the tree is labeled with symbols of syntactic relationship types. Over and above the valent quantifiers the SM language includes three limited primary quantifiers. An arbitrary sequence of restricted primary quantifiers and valent quantifiers is a prefix. An expression in the SM language is any expression of the narrow predicate calculus preceded (or not) by a prefix. During translation into the SM language the primary quantifier is substituted by the final one. Three quantifier order rules are given. The proper order of primary quantifiers in the prefix is secured by an ordering of the closure of relationships in the tree nodes and arranging primary and valent quantifiers in accordance with quantifier power. The general appearance of the rule for closure of "standard" syntactic relationships is given--predicative and semipredicative node.

SEMANTICS-DISOURSE: COMPREHENSION

Algorithm for the Construction of an Informational Notation for Text: Stage 1 (*Algoritm postroeniia informatsionnoi zapisi teksta. 1 etap*)

N. N. Leont'eva, and E. V. Uryson

Scientific-Technical Information, Collection, All-Union Institute of Scientific and Technical Information (Nauchno-tekhnicheskaiia informatsiia. Sbornik. Vse-soiuznyi Institut Nauchnoi i Tekhnicheskoi Informatsii," 1973, seriia 2), Series 2, No. 12: 3-13, 39: 1973.

An algorithm of automatic translation of a Russian sentence into an artificial language (AL) is described. The AL used is an artificial language with semantic properties similar to those of a natural one. A whole text is translated from the natural to information language in a series of transformations (stages). The first stage is translation of each individual sentence into the AL. That stage is termed the natural-to-semantic analysis and is described at length in the article. The principal goal of this stage is to establish correspondences between the units of the natural and the information languages. The subsequent transformation stages are briefly outlined, which eventually lead to the building of a complete informational notation of the whole text.

Elimination of Preposition-case Multiple Meaning Using a Thesaurus (*Ustranenie predlozhno-padexhnoi mnogoznachnosti na baze tezaurusa*)

V. I. Shabes, and V. V. Rozhkov

Statistics of Speech and Automated Analysis of Texts (Statistika rechi i avtomaticheskii analiz teksta) Science, Leningrad 231-244, 1972(1973).

A method for resolution of multiple meaning of lexical units belonging to a semantic class is proposed which makes use of a model of the corresponding segment of the semantic space formulated in terms of binary semantic features (SF). The model has the form of a tree (a figure is given) with two edges coming out of every node. The nodes are labelled with the symbols of SF, the SF standing at every two nodes *A* and *B* subordinated to a third, common node *C* are connected by the antonymy relationship and correlated with the SF at the node *C* as more concrete with more abstract. There are altogether 22 features each of which can assume the sign "+" or "-". The tree enables obtaining 2,352 strings of SF admissible for describing special relations. The SF which are to be assigned to a word are determined by considering language expressions containing that word. Each expression is assigned a certain string of SF, and the semantic characteristic of the word is obtained from the characteristics of expressions containing it by means of the operations of union and intersection of the corresponding sets of SF. The procedure is described in terms of set theory and illustrated by the preposition *from*.

SEMANTICS-DISOURSE: MEMORY

TORUS: A Step Toward Bridging the Gap between Data Bases and the Casual User

J. Mylopoulos, A. Borgida, P. Cohen, N. Roussopoulos, J. Tsotsos, and H. Wong
Department of Computer Science
University of Toronto
Toronto, Canada

Information Systems 2:49-64, 1976

TORUS (TORonto Understanding system) is a NL understanding system which serves as a front end to a data base management system which, in its current implementation, uses a semantic network to store information about a data base of student files. This knowledge is used to find the meaning of each input statement, to decide what action to take in the data base, and to select information for output. The system consists of a parser (which uses an ATN), a generator (which uses a version of the algorithm described by Simmons and Flocum, CACM 15: 891, Oct. 1972), a semantic network, a data base, a data base management system, and an interface. The semantic network contains nodes for *concepts*, *events*, and *characteristics*: such nodes represent generic ideas or instantiations of them. Generic nodes are organized in a superset-subset hierarchy with SUB arcs, while instantiations are linked to genera by E(xample of) arcs. Properties are inherited on SUB arcs. The net has provisions for the PART relation, 10 case arcs, and CHARACTERISTIC and VALUE which used in connection with characteristic nodes.

Analysis of Texts by Means of Speech Semantic Networks (*Analiz tekstov s primeneniem rechvykh semanticheskikh setei*)

F. F. Skorokhod'ko

In Mathematical Linguistics (Mat. lingvistika), T. 1. Kiev University: 97-103, 1973.

Two sentences are regarded as semantically associated, if the first of these sentences includes at least one relevant word which coincides with some word of the second sentence or is semantically connected with it by a generic or part-whole relationship. A speech semantic network (SSN) is defined as a two-dimensional graph representing the semantic relationships between various text units. Two types of SSN are distinguished, sentence-based and word-based. The possibility is discussed of using SSN to introduce quantitative parameters to describe the system of semantic relationships in text and its individual elements (words and sentences). The following experiment on text has been carried out: first, all nouns, adjectives and verbs each having one immediate relationship are eliminated from the text; then the words which have retained one relationship after the preceding operation are eliminated; the operation is repeated until only such words would remain in the text whose elimination would destroy at least two relationships at a time. Only a third part of the words remained but the resulting text was grammatically and semantically wellformed and conveyed the basic contents of the original text. Types of semantic structures of text are distinguished: chain, ring, piecelike (including sentence clusters, each connected with a majority of sentences of the same cluster) and monolithic (each sentence is connected with a majority of other sentences in the text). 3 refs.

SEMANTICS-DISCOURSE: TEXT GRAMMAR

Language Tools for Solving the Contradiction between the Multidimensionality of Contents and the Linear Organization of Text (*Nekotorye sposoby razresheniia iazykom protivorechiia mezhdru mnogomernost'iu plana sogerzhaniia i lineinoi organizatsiei tekstov*)

L. E. Pshenichnaia

Mathematical Linguistics (Mat. lingvistika), T. 1. Kiev University: 80-85, 1973.

The semantic plane of natural language--a complex situation expressed by a few sentences--is represented by a n -dimensional graph. There is a contradiction between meaning and form. Methods for fixation of the semantic relationships inside the linear sequence of text words are described. Within a sentence there is a tendency for increasing the complexity and the number of expressive means of the language serving to convey a semantic relationship in proportion to the increasing distance between words. The major means for identification of objects in different sentences to indicate the direct relationships of a word are as follows: repetition of the word; substitution by (local or contextual) synonym or a substitute word
10 refs.

Means for Naming the "Source" of Information in the Text (*Nekotorye Sposoby vyrazheniia v tekste "Istochnika" Informatsii*)

L. A. Starchenko

Linguistic Problems of the Functional Modeling of Speech (Lingvisticheskie problemy funktsional'nogo modelirovaniia rechevoi deitel'nosti.) Leningrad University: 140-144, 1973.

For automatic semantic analysis, text units are classified which signal the source of message (equalized to an individual sentence): (1) reference to the author's knowledge (I say that...); (2) reference to collective knowledge (According to quantum theory,...); (3) reference to a particular source (from the Boltzmann formula it is obvious that..., Rutherford supposed that...). Examples of text units of each type are given.

SEMANTICS-DISOURSE: TEXT GRAMMAR

Bracket Structures in Automatic Analysis (*Skobochnye konstruksii v sisteme avtomaticheskogo analiza*)

Antipenko E. S.

In Linguistic Problems of Functional Modeling of Speech (Lingvisticheskie problemy funktsional'nogo modelirovaniia rechevoi deitel'nosti). Leningrad University: 1-11, 1973.

For automatic syntactic analysis the functions are described of the bracket structures in English technical text on radiolocation (a file of 2,000 sentences). Round brackets are divided into unilateral and bilateral. Unilateral round brackets mark numbers and letters in enumerations. Bilateral round brackets single out: 1. a number or letter in enumeration (in the same way as unilateral ones, 2. a number or letter in a reference (usually after a formula) which functions as the identifying clause *the equation for R is: $R=(Pr)^n$ (1)*; 3. a word (in particular, an abbreviation); 4. a sentence or a pair of sentences; and 5. a phrase. Sentences closed in brackets should be singled out as a segment. There are such phrases which have the same function as the sentence and should be segmented correspondingly. The formal feature of this type of phrases is the verb *to see* in the imperative form: *The radar modulator(see Fig. 3) consists of...* or a personal verb with an *S* without a preposition to the left: *The figure shows where the target is located (the center of the target is...)* Examples of phrases that are not independent segments are noun phrases functioning as sentences and participle phrases. Furthermore, round brackets as well as square brackets and braces are used in mathematical expressions.

LINGUISTICS: METHODS: STATISTICAL

On the System Character of Linguistic Statistics (*O sistemnom kharaktere lingvostatistiki*)**P. M. Alekseev***Theory of Language and Engineering Linguistics (Teoriia vazyka i inzh. lingvistika), Leningrad: 66-73, 1973.*

The notion of "system" is considered as interpreted by some Soviet logicians. The set of linguistic objects is stated to be a system (at least "badly organized" or "diffuse" as defined by V. V. Naïimov). The linguostatic studies conducted by the "Speech Statistics" research group emphatically do conform to the adopted approach to system analysis. The group has pioneered in comprehensive research on various linguistic systems (in particular, owing to a uniform plan and method, some 80 frequency dictionaries of various language subsets already compiled can be integrated in various combinations, etc.) The author regrets that as yet many linguists have failed to master the method of statistical analysis. The fact that there is no adequate textbook is to blame for that. 12 refs.

ARTIFICIAL INTELLIGENCE

Artificial Intelligence Meets Natural Stupidity**Drew McDermott***M.I.T. Artificial Intelligence Laboratory
Cambridge, MA 02139**SIGART Newsletter 57: 4-9, April 1976*

Three common types of AI hubris are offered for consideration. 1) *Wishful mnemonics*: Calling the main loop of a program "UNDERSTAND" makes it too easy to beg the important theoretical question behind an "understanding" program: What and how and in what sense does it understand? Wishful labeling of links leads to the same type of question-begging (the ubiquitous IS-A link is discussed). 2) *Unnatural language*. AI workers tend both to oversimplify and overglorify the problems of NL processing. NL work deserves more attention from a theoretical point of view before we rush off and throw together "NL" interfaces to programs with inadequate depth. 3) *"**Only a Preliminary Version of the Program was Actually Implemented"*: That is, don't count your chickens before they're hatched. It will just cause trouble for the next guy, who may really need chickens, not a pile of rotting eggs.

**On Context-Sensitive Substitutions
(Kontekstno-chuvstvitel'nykh podstanovkakh)**

B. E. Kats, and I. M. O. Reithort

Scientific-Technical Information, Collection, All-Union Institute of Scientific and Technical Information, Series 2, No. 8: 38-39, 45, 1973. ("Nauchno-tekhnicheskaya informatsiya. Sbornik. Vse-soiuznyi institut nauchnoi i tekhnicheskoi informatsii," 1973, seria 2,)

Symbol-by-language replacement which is performed by a (context-free) substitution does not depend on the context of the symbol. A generalization of context-free substitution is introduced so that the replacement becomes associated with the context. It is proved that: (1) context-sensitive substitutions have, in a sense, a wider scope than context-free substitutions; (2) many classes of languages are closed under context-sensitive substitutions. 7 refs.

MATHEMATICS

**Objectives of Mathematical Linguistics
(zagachakh matematicheskoi lingvistiki)**

G. P. Bagrinovskaya, and A. A. Liapunov

In A Few Questions on Theoretical Cybernetics and Algorithms of Programming, Novosibirsk: 95-110, 1971. O nekotorykh voprosakh teoreticheskikh kibernetik i algoritmov programirovaniya, Novosibirsk 1971

Linguistic problems requiring construction of mathematical language models are discussed: 1. Identification and classification of the elementary linguistic objects; 2. Study of the types of combinability of various classes of morphemes. Mathematical problems facing mathematical linguistics are discussed: (a) mathematical description of the set of texts of a given language; study of the invariance conditions of a language described with respect to certain transformations; construction of an algorithm; (b) comparison of sets of texts as generated by certain strictly formal models; (c) construction of models for translation of languages; and (d) statistical studies of languages. Mathematical linguistics is intrinsically connected with theoretical linguistics and is its natural component. Mathematical linguistics deals with mathematical models of linguistic phenomena. The concept of "modelling" is explained, discussing a range of associated notions such as simulation, automation, algorithmization and computer implementation. 7 refs.

Overview of the Hearsay Speech Understanding Research

Lee D. Erman

Computer Science Department, Carnegie-Mellon University, Pittsburgh, PA 15213

SIGART Newsletter 56:9-16, February 1976

Several independent knowledge sources (KSs) interact in a hypothesis-and-test paradigm and are interfaced through a dynamic global data structure, the *blackboard*. Hearsay I: KSs are activated in a lockstep sequence of three phases per cycle: *pool, hypothesize, test*. All KSs are activated at each phase and the next phase doesn't start until all KSs have completed the present one. A mediator module maintains the blackboard, calculates combined ratings from ratings assigned to hypotheses by the individual KSs and decides when to stop and accept a solution (or to give up). The currently highest rated hypothesis is the one used as the context for the next cycle. Hearsay II: The blackboard has been extended and generalized to allow a) the representation of all levels of information (acoustic phonetic, syllabic) in addition to the lexical and sentence level of Hearsay I, and b) the explicit representation of relationships among hypothesis. The overall control strategy is considerably more flexible than that for Hearsay I.

DOCUMENTATION: INDEXING

Subject Indexes to VINITI's Abstract Journals: Efficiency Measurement (*Otsenka effektivnosti predmetnykh ukazatelet k RZH VINITI*)

V. V. Bel'tsova, B. B. Mokshantsev, N. M. Sagalevich, and A. V. Fokina

Questions on the Perfected System of Informational Publication, Moscow: 79-98, 1973.
Voprosy sovershenoi sistemy informatsionnoi izdani M., 1973

A method for information efficiency measurement of subject indexes to three separate series of VINITI's abstract journals is described, and the results of its evaluation are presented. The information efficiency, γ , is shown to be convenient for evaluating the efficiency of use of the subject indexes. The value of γ_1 can be used as a measure of perfection of the subject index vocabulary and, by that, as an objective basis for further improvement of the vocabulary. Representative data for the entire set of subject indexes can be obtained only from a multifactor experiment of the 2^n type, where n is the number of factors for the whole set of VINITI's abstract journal subject indexes. 8 refs.

Specifics of Automatic Translation in Artificial Languages (with Special Reference to Chemical Nomenclature) (Osobennosti avtomaticheskogo perevoda v iskusstvennykh iazykakh (na primere iazykov khimicheskoi nomenklatury))

L. V. Knorina

Scientific-Technical Information, Collection, All-Union Institute of Scientific and Technical Information, Series 2, No. 8: 30-33, 46, 1973. ("Nauchno-tekhnicheskaiia informatsiia. Sbornik. Vse-soiuznyi institut nauchnoi i tekhnicheskoi informatsii" 1973, serii 2,)

The characteristics of automatic translation in artificial languages are illustrated by an algorithm for translation of names of a group of chemical compounds from the Hunch Wiedmann nomenclature to the a-nomenclature. An algorithm is described; its essential function is building the patterns of the structures corresponding to the names of compounds to be translated. The need is stressed for distinguishing between translation and interpretation in algorithmizing artificial language translation.

TRANSLATION

Machine Translation: Aspects of Methodology (Otnosiashchikhsia k mashinnomu perevodu)

G. P. Bagrinovskaia, O. S. Gulagina, and A. A. Liapunov

In A Few Questions on Theoretical Cybernetics and Algorithms of Programming, Novosibirsk: 67-94, 1971. ("O nekotorykh voprosov teoreticheskikh kibernetik i algoritmakh programmur", Novosibirsk 1971)

The following issues are discussed: 1. Place of machine translation in cybernetics. 2. Goal of machine translation research--creation of an operational machine translation system capable of accelerating the translation process as a whole and making it less expensive. 3. Organization of the work on machine translation development. 4. Problems to be faced in developing a machine translation algorithm: choice of text, decisions as to the general type of algorithm, elaboration of the rules of an algorithm, compilation of the dictionary, etc. 5. Programming. 6. Automation of work in machine translation algorithm construction. 7. Machine translation and allied areas. 8. Universal vs. specialized machine translation algorithms. 9. Role of scanning. 19 refs.

TRANSLATION

**Analysis of Formulas in Automatic Translation of Scientific and Technical Text
(Analiz formul pri avtomaticheskom perevode nauchno-tekhnicheskikh tekstov)**

G. M. Kirillova

Linguistic Problems of Functional Modeling of Speech (Lingvisticheskie problemy funktsional'nogo modelirovaniia rechevoi deiatel'nosti), Leningrad University: 37-43, 1973.

The paper discusses the processing of mathematical formulas during automatic translation (with special reference to English text in radiolocation--2,000 sentences). Formulas are subdivided by their form into "pure" (those containing only mathematical symbols and characters) and "mixed" (formulas containing words whose stems can be found in the dictionary: *f=local microwave reference frequency*). By syntactic structure, formulas are broken down into formant (those including one of the predicative signs, and nonformant (containing no predicative signs). Signs of arithmetic operations are regarded as nonpredicative. Nonformant formulas function in the sentence as a word: an ordinal numeral if followed by the suffix *th* (*(n-1 th)*) or a noun (in the remaining cases: *if $F(Q_1)$ is less than...*). A formant formula *F* can be isolated from the sentence as an individual sentence in its own right except in three cases: 1. there is the verb *to let* to the left of *F*; 2. *F* stands to the left or to the right of the personal verb; and 3. *F* stands after a preposition. In these three cases *F* is treated as a noun.

TRANSLATION

**The Tasks of the Transformational Stage in an Automatic Translation System
(Zadachi etapa preobrazovaniia v sisteme AP)**

B. M. Leikina

Linguistic Problems of Functional Modeling of Speech (Lingvisticheskie problemy funktsional'nogo modelirovaniia rechevoi deiatel'nosti), Leningrad University: 33-37, 1973.

A tentative English-Russian translation system is based on the principle of independent analysis and synthesis and correspondingly includes the intermediate stage of transformation. Transformation I is responsible for the transition from the syntactic structure (dependency tree) of the input sentence to the structure of the sentence in the intermediate language (IL). Transformation II (which is a mirror image of transformation I) modifies the latter structure in keeping with the specifics of the output language and produces information required for morphological synthesis of a Russian sentence. The grammars of transformation are a set of rules applied in a fixed order. Twenty two English sentences and their translations as obtained in an experimental setting of the algorithm are presented.

On the Machine Linguistic "Sign" (*O mashinnom lingvisticheskom "znake"*)**A. D. Nakhimovskii**

Theory of Language and Engineering Linguistics (Teoriia yazyka i inzh. lingvistika), Leningrad: 160-166, 1973.

A machine sign (MS) is the set of input linguistic units (the signifier of MS) and an output unit(s) (denotation of MS) which has been put into correspondence with it in the course of text processing. The referent of an MS is the machine document(s) which establishes the correspondence between a unit in the input language and a unit in the output language. The translation program delivers the denotation of MS. The language of machine instructions turns out to be thereby the metalanguage of description of machine semantics (and, moreover, it is a completely terminological metalanguage). During automatic indexing the signifier of MS is the text to be indexed as a whole and its denotation is the indexation, that is, the subject field. Descriptors singled out in the text and matched against the standard set of descriptors of the field constitute the distinctive features of MS and all other text elements are its integral features. This language is particularly abundant in synonyms, but sets of synonyms constitute equivalence classes, as indexing is performed in a unique way.

Study of the Translation Ambiguities of English Nouns with Special Reference to Parallel Texts in Science and Engineering (*Nekotorye rezul'taty issledovaniia perednoi neoanachnosti angliiskikh sushchestvitel'nykh na Materiale nauchno-tekhnicheskikh parallel'nykh tekstov*)

E. A. Pustovoitova

Theory of Language and Engineering Linguistics (Teoriia iazyka i inzh. lingvistiki), Leningrad: 103-112, 1973.

A study of translation ambiguities (variability of translation units) of English nouns was conducted in two stages. 1. From the journal *Electronics*, a sample totalling 24,243 English running words was extracted; it contained 3,622 different text words. Parallel texts from *Electronics* were used to compile an English-Russian alphabetic word frequency list, which fixed for every English word form its frequency, all translation Russian word forms, and separately the lexemes. 2. Textual ambiguities of the most frequent and equivocal nouns were compared with their dictionary ambiguity as determined by the English-Russian dictionary (general, technical, and specialized) and Roget's thesaurus, where the number of meanings was measured by the number of associative groups. The comparison indicated that for the bulk of nouns there is a natural decrease of ambiguity as the scope of the dictionary is narrowed down and as one passes over to text (e. g., the words *bit, line, character, light*, etc.). A minority of nouns in that case increased their ambiguity (e. g., *signal, transistor, voltage, diode, resistor*, etc.). Some nouns (*circuit, number, source, data*) lowered ambiguity on passing over from the general to the technical dictionary, while it was again increased in specialized dictionaries and in text. These deviations from the natural tendency to decrease of ambiguities are explained by inadequacies of the dictionaries.

TRANSLATION

On Machine Translation of Natural Languages (*O mashinnom perevode tekstov na estestvennykh iazykakh*)

O. S. Kulagina

Problems of Cybernetics (Problemy Kibernetiki) 27:33-45, 1973.

Theoretical research has led to interesting results in the field of mathematical linguistics and determined a wide scope of problems that must be solved for development of MT systems. The MT systems that were developed in the first years following the Georgetown experiment are briefly outlined. In the second-generation systems (developed in the 60s) translation involves the following steps: dictionary lookup, phrase processing (idiomatic phrases), and morphological analysis; syntactical analysis using transformation rules; syntactical synthesis; and morphological synthesis. These stages are exemplified by the system of French-Russian MT developed at the Institute of Applied Mathematics of the USSR Academy of Sciences. Third-generation systems are now under development. A transition to them would mean improvement of analysis (deepening, transition to deep syntactic structures, and widening, going beyond the limits of the sentence) and of synthesis (ensuring greater flexibility and idiomaticity), as well as refinement of mathematical formalisms.

TRANSLATION

On the System of French-Russian Machine translation (FR-II) (O Sisteme frantsuzska-russkogo mashinogo perevode FR-II)

O. S. Kulagina

Problems of Cybernetics (Problemy kibernetiki) 27:47-61, 1973.

FR-II was developed at the Institute of Applied Mathematics of the USSR Academy of Sciences and is intended for the translation of mathematical texts. The dictionary has three parts (French and Russian parts of the dictionary of stems, and the dictionary of phrases); it includes some 1,300 French words, their Russian equivalents, and about 220 phrases. The system translates one sentence at a time in the following steps: 1. dictionary lookup; 2. processing of phrases (word combinations which cannot be translated word-by-word); 3. morphological analysis; 4. syntactic analysis (using a multiple algorithm of the filter type); 5. processing of prepositions; 6. transformations; 7. syntactical synthesis; 8. morphological synthesis. FR-II has been implemented on a BESM-4 computer; an example translation (or a 15 sentence fragment) is given.

TRANSLATION

Automatic Translation of English Civil Engineering Texts (Avtomaticheskii perevod angliksikh stroitel'nykh tekstov)

a. D. Borisevich

Linguostatistics and the Automated analysis of Texts (Lingvostatistika i avtomat. analiz tekstov), Minsk: 279-285, 1973.

The completion of a project on a word list for the English-Russian automatic translation system in civil engineering is reported. The word list is to include only special terms; the common vocabulary was selected earlier and is considered to be applicable to all language subsets. The project comprised three stages: (1) selection of English terms; (2) finding Russian equivalents for these terms (resp. compilation of the Russian list of terms); and (3) development of the system for coding grammatical and lexico-grammatical information. The English part of the dictionary includes some 7,000 word forms and 7,500 phrases. To test the exhaustiveness of the list and the efficiency of the lexico-grammatical information, as well as of the rules for transition to Russian equivalents, a computer experiment has been carried out: the Minsk-22 computer produced correct equivalents for 97 per cent of word forms and 96 per cent of phrases.

TRANSLATION

Solving the Problem of Interlanguage Idiomaticity in Automatic Text Processing (*Razreshenie problemy mexh'iazykovoï idiomatichnosti pri avtomaticheskoi pererabotke teksta*)

A. D. Borisevich

Linguostatistics and the Automated Analysis of Texts (Lingvostatistika i avtomat. analiz tekstov), Minsk: 266-278, 1973 2("," Minsk)

The paper reports work on a dictionary of idioms (I) intended for automatic text processing, primarily automatic translation (English); I is defined as frequent phrases of the input language that cannot be translated in a regular way either using the existing automatic dictionaries of word forms, or with reference to morphological or syntactic algorithms. It is proposed to compile the dictionary of common I to be supplemented with special dictionaries of I for each specific subset of language (electronics, building materials, etc.). The criteria for selection of I are briefly described, the main being the occurrence frequency. For searching I a frequency list of phrases of different lengths were used (mostly triads). Translation of I into Russian was a problem in itself. Essential for adequate translation was the choice of "reference" word in the idiomatic phrase. Criteria for this choice are discussed. As a rule, the "reference" word had the least frequency as compared with other components of the idiomatic phrase. The preliminary list of I was tested on 12 subsets of language (2 million running words). All I from the list have been found in the text samples with a frequency of 1 to 80.

SOCIAL-BEHAVIORAL SCIENCE: PSYCHOLOGY

Analysis of Some Numerical Characteristics of Statistical Speech Studies in Mental Pathology (*Analiz nekotorykh chislobykh kharakteristik statisticheskogo issledovaniia rechi pri patologii myshleniia*)

B. M. Leikina, M. I. Otkipshchikova, F. I. Sluchevskii, G. S. Tseitin, and B. A. Shoherbatov

Linguistic Problems of the Functional Modeling of Speech (Lingvisticheskie problemy funktsional'nogo modelirovaniia rechevoi deitel'nosti), Leningrad University: 93-102, 1973.

Numerical characteristics of studies of schizophrenic and epileptic speech are given which were obtained by computer processing of statistical samples of pathological speech. Only syntactic features are considered. For each feature the mean values of a given parameter are given as found for schizophrenic and epileptic patients respectively and the difference measure (D. M.) between the two. The greatest deviation from the random distribution was observed for such features as the average length of anaphoric relationship (D. M.=28.8); the average size of a syntactic group (d. M.=18.9); the depth of a node in a tree--the average distance from a node to the apex (d. M.=18.9); the average length of relationship (D. M.=15.8); the average length of an arrow (D. M.=14.5); and the analogue of Yngve's depth (D. M.=12.3). The coefficients of other features varied within random deviations (for example, the proportions of individual parts of speech).

Association for Literary and Linguistic Computing: Bulletin

Volume 4, No. 2, 1976

CONTENTS

- GUEST EDITORIAL: TEXTEDITION - TEXTCORPUS - EDV,
ZUM PROBLEM DER PHILOLOGISCHEN
GRUNDLAGEN FUER DIE LINGUISTISCHE DATENVERARBEITUNG, F. V. Spechtler . 95
- ON LEXICOGRAPHICAL COMPUTING: SOME ASPECTS OF THE WORK FOR A
MEXICAN SPANISH DICTIONARY, L. F. Lara 97
- MACHINE TRANSLATION: PAST, PRESENT, FUTURE, S. -C. Loh 105
- L'ENREGISTREMENT INFORMATIQUE DE DOCUMENTS NON-LITTERAIRES REDIGES
DANS UNE LANGUE ANCIENNE: OBSERVATIONS ET SUGGESTIONS, G. P. Zarri . 115
- COMPUTER ASSISTANCE IN DECIDING THE AUTHORSHIP OF AN INDIAN
POEM, S. K. Havanur 125
- ZUM PROJEKT VERSKONKORDANZ DER MINNESANG-HS. B. U. Muller 131
- STATISTICAL DATA FROM THE DICTIONARY OF THE CZECH LITERARY
LANGUAGE, Jitka Stindlova 135
- A LANGUAGE FOR LITERARY DATA PROCESSING, I - THE CHOICE OF A
LANGUAGE, A. M. Addyman 146
- THE SECOND INTERNATIONAL CONFERENCE ON COMPUTERS IN
THE HUMANITIES, A. Aebi, R. Hirschmann, R. Moak 152
- COMPTE-RENDU DE LA TABLE RONDE SUR L'UTILISATION DE
L'INFORMATIQUE POUR L'EXPLOITATION DES DOCUMENTS TEXTUELS MEDIEVAUX:
Rome. 20-22 mai 1975, Lucie Fossier. 155

HUMANITIES

Association for Literary and Linguistic Computing: Bulletin

Volume 4, No. 3, 1976

CONTENTS

GUEST EDITORIAL, K. Hyldgaard-Jensen	189
<u>ENCODING</u> LITERARY TEXTS: SOME OBSERVATIONS, J. B. Smith	190
PREPARING CONCORDANCES TO OLD ICELANDIC LEGAL TEXTS, H. Fix	190
MASCHINELLE DATENVERARBEITUNG IM DIENST HISTORISCHER VALENZSYNTAX, GEZEIGT AM BEISPIEL EINES LUTHER-TEXTES, J. Korhonen	206
ON LITERARY CONCORDANCES: AN EARLY VIEW, T. H. Howard-Hill	215
ON GENERATIVE GRAMMAR ANALYSIS: RECURSION IN SNOBOL, J. D. Blackburn	221
PROCESSUS INTERACTIF DE SAISIE DE VISUALISATION ET D'IMPRESSION DES DONNEES SEMIQUES ANCIENNES ET MODERNES, G. E. Weil, G. Opsomer	223
A LANGUAGE FOR LITERARY DATA PROCESSING, II - SIMPLE ALGOL 68, A M. Addyman	238
THE USE OF THE COMPUTER IN LINGUISTIC AND LITERARY RESEARCH: Fourth International Symposium, Oxford, 5-9 April 1976, A. R. Pester	245

Computational Linguistics and the Humanist

William Benzon

*Department of English
State University of New York
Buffalo, New York 14260*

David G. Hays

*Department of Linguistics
State University of New York
Buffalo, New York 14260*

Computers and the Humanities 10: 265-74, 1976

We start with a consideration of the nature of computational linguistics followed by a consideration of how to formulate problems in the analysis of literary texts so that the techniques can be fruitfully employed. Rather than adapting one's conception of literature to what one can do with a computer, one must adapt one's computational model to the demands of literary analysis. If you wish to analyze literary texts, then you need a theory of human literary performance with which to conduct the analysis. Next comes a review of current work in semantics and discourse analysis (Wilks, Schank, Norman, Kintsch, etc.) and in speech understanding (SPEECHLIS, VDMS, HEARSAY). Finally, a brief sketch shows how one can start a long-term Shakespeare project with the analysis of a sonnet or two, selected passages from the plays, and plot structures (abstracted from their detailed realization in the drama) and, as our knowledge of literature and of NL computing systems develops, aim toward an analysis of a complete play (which would probably require a computer whose architecture is organized as a cognitive network).

HUMANITIES: ANALYSIS

Intersentence Relationships in various Styles of Written Language (*Mezhhfrazoyve sviazi vrazlichnykh stilliakh pis mennogo iazyka*)

O. A. Borodachenko

Mathematical Linguistics (Mat lingvistika), T. 1. Kiev University: 14-25, 1973

More than a hundred different types of intersentence relationships have been established, which are divided into six classes: 1. connective elements; 2. repetition of lexical forms; 3. common stem words; 4. substitutions; 5. structural similarity and dependence of sentences; and 6. synonymic relationships. Fifteen most common types of connection have been singled out. Distant reactance of each type of relationship has been studied. Fiction is characterized by more complex relationships implemented using more formal means simultaneously. The peculiar features of intersentence relationships in English, as compared with Russian, are discussed. Other possible approaches to the study of intersentence relationships are considered; context semantic fields, "transformational" approach, etc. 10 refs.

Understanding Understanding Poetry

Stanley R. Petrick

*IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598*

IBM Research Report RC 6270, 9 pp., September 10, 1976

First, a brief review of the state of the art in understanding systems suitable for answering English questions relative to a given data base. Some of the difficulties discussed are equally applicable to the understanding of poetry, and we illustrate them by considering a few problems encountered in understanding a stanza of Donne's *A Valediction: Forbidding Mourning*. And, as difficult as it is to characterize the grammar of standard prose the job of appending appropriate modifications to yield a grammar of poetry is likely to be harder. The problem is even worse with respect to world knowledge: what sort of world knowledge is necessary for the understanding of poetry? And how does one determine whether or not a computer has understood a poem? Finally, though current poetry generation programs *seem* more impressive than any current NL understanding programs, the poetry they produce doesn't stand up under skilled critical scrutiny. This suggests that a good poetry generation program will have to include the essentials of an adequate model of poetry understanding.

HUMANITIES: ANALYSIS

Sentence Patterns in Japanese Literature

Akio Tanaka

Keiryō kokugogaku Mathematical Linguistics 65: 8-22, 57, 1973

The frequency of occurrence of noun groups having various types of case forms are counted (case particles *wa, no, ga, o, to, mo*) as occurring at the beginning of the sentence, as well as the frequency of verbal forms and the forms of predicative adjectives (*masita, masen, mas, desita, masen desita, ta no des, da, da se, no des, ta, nakatta, datta, etc.*) at the end of the sentence. Frequency data are given on sentences with various syntactic structures. At the beginning of the sentence the highest frequency is possessed by the structure "noun + *wa*", then stands "noun + *no*" (in particular, in newspaper texts the former structure occurs in same 24 per cent of the sentences). At the end of the sentence the highest frequency is possessed by the verbal form with the *masita* suffix and also with the *ta* suffix. On the basis of these data a probabilistic model of the beginning of the Japanese sentence has been constructed (different models for the texts analyzed), as also a probabilistic model of the end of the sentence. By merging the two models, a quasimodel for Japanese simple sentence is obtained (separately for each text).

Aspects of Expression in Sentence-ends**Takashi Ichikawa***Keiryō.kokugogaku, Mathematical Linguistics 65: 1-8, 57, 1973*

At the end of the Japanese sentence a verb or predicative adjective normally occurs (a noun can occur only in colloquial speech, in sentences with an elliptic predicate and in some standard document forms). At sentence ends there are such forms as a verbal stem + *ta*; a stem of the predicative adjective + *ta*; a predicative + link verb *datta* (*de atta*); a verbal stem + *te ita*, etc. Proceeding from the grammatical form of verb at the end of the sentence, all sentences of modern Japanese are broken down into 12 types. Using appropriate formulas the uniformity of the distribution in text of various types of sentences and the distribution of sentences of one type from the viewpoint of their relative positions can be measured. Three works by the 20th century Japanese writer Natsume Soseki have been analyzed in these terms.