NBS

Technical Note 258

A PREREQUISITE
TO THE UTILITY OF MICROGRAMMARS

WILLIAM C. WATT

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

THE NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards is a principal focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. Its responsibilities include development and maintenance of the national standards of measurement, and the provisions of means for making measurements consistent with those standards; determination of physical constants and properties of materials; development of methods for testing materials, mechanisms, and structures, and making such tests as may be necessary, particularly for government agencies; cooperation in the establishment of standard practices for incorporation in codes and specifications; advisory service to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; assistance to industry, business, and consumers in the development and acceptance of commercial standards and simplified trade practice recommendations; administration of programs in cooperation with United States business groups and standards organizations for the development of international standards of practice; and maintenance of a clearinghouse for the collection and dissemination of scientific, technical, and engineering information. The scope of the Bureau's activities is suggested in the following listing of its four Institutes and their organizational units.

Institute for Basic Standards. Electricity. Metrology. Heat. Radiation Physics. Mechanics. Applied Mathematics. Atomic Physics. Physical Chemistry. Laboratory Astrophysics.* Radio Standards Laboratory: Radio Standards Physics; Radio Standards Engineering.** Office of Standard Reference Data.

Institute for Materials Research. Analytical Chemistry. Polymers. Metallurgy. Inorganic Materials. Reactor Radiations. Cryogenics.** Office of Standard Reference Materials.

Central Radio Propagation Laboratory. ** Ionosphere Research and Propagation. Troposphere and Space Telecommunications. Radio Systems. Upper Atmosphere and Space Physics.

Institute for Applied Technology. Textiles and Apparel Technology Center. Building Research. Industrial Equipment. Information Technology. Performance Test Development. Instrumentation. Transport Systems. Office of Technical Services. Office of Weights and Measures. Office of Engineering Standards. Office of Industrial Services.

---

* NBS Group, Joint Institute for Laboratory Astrophysics at the University of Colorado.
** Located at Boulder, Colorado.

NATIONAL BUREAU OF STANDARDS

Technical Note 258

A PREREQUISITE
TO THE UTILITY OF MICROGRAMMARS

William C. Watt

CONTENTS

A PREREQUISITE TO THE UTILITY OF MICROGRAMMARS

William C. Watt

This paper takes up the question of a hitherto-ignored obstacle to the useful functioning of microgrammars in artificial intelligence systems. This obstacle consists of the difficulty of "staying within" the microgrammar in man-machine communication, a condition rooted in the fact that microgrammars produce a "language" which consists entirely of English sentences, but of only some English sentences: and it is hard or even impossible for the microgrammar-user to remember which sentences he is allowed to use. Besides raising this problem, and studying it in some detail, I indicate what steps may be taken to overcome it; these are such steps as lend the microgrammar more "extrapolative symmetry".

0.   Introduction[1]

Many people whose research makes essential use of large digital computers can 'converse' freely with their machines without feeling hampered by their being unable to converse in their native tongue. For example, mathematicians presumably feel little need to express in English their instructions for high-speed computation, since ALGOL serves this purpose quite adequately[2]; nor would they react with anything but annoyance if the computer were given to responding in English, rather than in mathematical expressions and organizations of such expressions. However, in other situations, the unavailability of

---

2   I will continue to use ALGOL to exemplify the large multi-purpose programming languages; obviously another, such as FORTRAN, would have served about as well. In the same fashion, I will continue to use English as my example of a natural language.

English as a man-machine language is felt rather keenly, for either or both of two reasons[3]. First, there are many instructions (and queries) which are better expressed in English than in ALGOL, and many answers more suitably phrased in English than in the form of data-structures. And secondly, those who want to communicate with the machine may not be conversant with any of the machine languages, and may (as is generally the case) be at a point in their professional careers where taking off the time required to become proficient in such a language would be out of the question.

Machine languages can be extended, of course: ALGOL could be given the power to express more than it now does, and could perhaps be progressively extended so that in the end it would be able to express anything expressible in English. However, such a procedure would have the secondary effect of making ALGOL more and more difficult to learn: we can lessen the first of ALGOL's liabilities (its deficiency in expressive power) only at the cost of greatly increasing the second (its relative 'distance' from the user).

There are good reasons, then, for wishing that English were available as a man-machine language. But for English to serve this function it must first have been described by a quasi-complete grammar ((qcg)), roughly analogous to the syntax of ALGOL; and this qcg, moreover, must be in a model to which computers are accomodated. Despite intensive efforts at more than one linguistic center[4] this first condition has not yet been met, though it may be in the relatively near future[5]; enough is already known about English, however, to indicate

---

3 I will argue in another paper that this lack should be felt even more keenly, in these 'other situations', for more compelling reasons which as yet are not widely recognized.

4 Chiefly at the University of Pennsylvania, at Harvard, and at M.I.T.; see especially ((3)) for results of the research at Penn. Numbers in double parentheses (( )) refer to items in the appended list of References.

5 On the other hand, an 'English grammar' in the most inclusive sense of this term--a device which simulates speaker-behavior in its linguistic aspects (if these can be delimited) is hardly even contemplated at present, except to be shelved as being forbiddingly, perhaps impossibly, difficult.

that meeting the second condition --- getting a qcg into the computer --- may be extremely difficult, for English in its 'entirety' (probably any natural language) requires a model of very powerful capacity[6].

Inevitably, then, attention has occasionally turned toward the practicability of constructing useful computer grammars for <u>portions</u> of English. There is no doubt that such partial grammars can be written --- such "microgrammars", as I will call them --- in fact several, of varying sizes, are already in existence[7]. I believe that it is not generally realized, however, that there is a formidable obstacle in the way of these microgrammars' being of real use to a synthetic intelligence system.

This paper examines that obstacle and indicates how it may be overcome.

1. <u>Extrapolation, and the Avoidance of Its Pitfalls.</u>

A Microgrammar allows its user to employ English sentences in communicating with the computer; but, by definition, it allows him to employ only <u>certain</u> English sentences. The user is invited to speak in his native tongue, but he is also enjoined to choose his sentences carefully lest he express something which the microgrammar is powerless

---

6   This question has been widely treated in the literature, most concisely in ((2)). I place the word "entirety" in quotes because the boundary-line between 'English' and 'non-English' is by no means sharply defined. Even when a large number of utterances have been satisfactorily ranked by decreasing grammaticality, it is not easy to set that threshold below which the utterances are so ungrammatical as to be unEnglish.

7   Of those I am acquainted with the most highly developed are the large microgrammar written by Jane J. Robinson at the RAND Corporation ((9)), and the smaller PLACEBO IV, written by the present author at the National Bureau of Standards ((11)) and ((12)). It should be emphasized that e.g. ((3)) covers much more English than either Robinson's grammar or PLACEBO IV; but ((3)) shows not a microgrammar but hopefully a massive segment of a qcg. The difference between these two types of algorithm is made reasonably clear in ((11)).

3

to analyze: which therefore the machine is powerless to act on. Clearly then the user must somehow become familiar with the set of allowable utterances.

To draw a comparison, let us suppose that an American archeologist is introduced to a French colleague who he is told "speaks a little English", "enough to conduct a conversation about archeology". Realizing that he and the Frenchman share only a very small portion of English, and wanting to waste as little time as possible, the American will want to learn quickly what the limits of this shared portion are. If he is sensible, rather than launching into a discussion and taking his chances sentence by sentence that he will overstep the portion's boundaries, the American may go about his task systematically. Restricting himself first of all strictly to archeology, so as to limit vocabulary, he may hazard one or two simple sentences; finding these understood he may chance a few more along the same lines; and thus by extending little by little the bounds of discourse, he may succeed in gradually marking off an area of English within which he and the Frenchman can converse, with but very few wasteful oversteps. The intuitive 'system' he will have made use of is one which it would be interesting to know more about; for the moment at least let us be content with calling it one of 'rough extrapolation', a process exemplified by the reasoning: "x and y are English sentences and are similar, and x was within the bounds of the shared portion of English, therefore y must be".

To illustrate the results of such extrapolation, let us glance at the first sentences which such a hypothetical American might produce when asked to summarize his own views on Maya-Toltec interactions.

1. "Teotihuacan strongly influenced the Guatemalan highlands."
2. "The Toltecs dominated the Yucatecan Maya."
3. "The dominance of the Toltecs decreased during the Mayapan period."
4. "This decrease coincided with the flowering of the Guatemalan City-States, Iximche for example."

5. "Whether or not this coincidence was meaningful, is open to question."

Note that, over this series of sentences, there is a progression toward greater complexity: of structure and of relation to the preceding sentences[8]. The first sentence typifies the English Subject-Verb-Object sentence, where both Subject and Object are nominals. The second sentence has the same structure. In the third, the Subject is a nominalization of the verb of the second sentence; the Subject of the fourth sentence is a nominalization of the Verb of the third. In the fifth sentence, the Subject is in part a nominalization of the Verb of the fourth, but the subject has been complicated by the "whether or not" construction. (Needless to say, this rough description of these intersentential relations is not meant to be taken very seriously.)

It is fair to say, I think, that our hypothetical archeologist did not do a bad job. He adhered to as simple a vocabulary as he could[9], and he kept his syntax 'simple' in some untutored meaning of that term. I think it almost self-evident, however, that if he continued the process much further, he would soon run afoul of the Frenchman's limitations. In fact, he may have already done so in sentence five: the Frenchman may be unable to parse the "whether or not..." Subject, and may thus fail to understand this sentence[10]. The American in this

---

8   Of course I do not mean to imply that, in an actual conversation under the stated conditions, any such progression would appear so dramatically, in the span of five sentences. Still less do I claim that in any conversation complexity could or would increase unendingly. I mean only to exemplify a process which we might well expect to find in use under the stated circumstances. If this rather casual example be taken as an hypothesis, it should not be a very hard one to test.

9   He can also be said to have profited from his knowledge of French, in using the cognate "decrease" rather than the equally natural "wane".

10   The Frenchman's total inability to understand a sentence which he can parse only in part, will be questioned below.

case would be forced to backtrack, to try a simpler way of conveying his thought. There is no good reason to believe, in fact, that the American will ever entirely cease to overstep the boundaries of the Frenchman's portion of English, unless he restricts himself to so small a vocabulary, and so scrawny a syntax, as to render quite easy his observance of the limits, but almost impossible his transmission of information.

A microgrammar, too, 'understands' only a small set of English sentences; and in this respect the user of a microgrammar is in the same fix as the American archeologist sketched above. He must watch himself carefully lest he trespass on forbidden territory, while at the same time uttering new sentences in order to get his message across. Naturally if the microgrammar is extremely small the user can simply memorize the list of its allowable sentences; if it is moderately small he may succeed in memorizing its grammatical rules, and perhaps (this is less likely) in composing sentences which accord with those rules[11]. In fact, however, a microgrammar of such limited scope could scarcely be useful as a vehicle of expression: it would probably be no more useful as such than the sort of linguistic apparatus one can acquire by thumbing through a tourist phrasebook[12]. Short of memorizing the sentences and/or the rules, the user must either gradually _infer_ the boundaries of allowable speech, as our archeologist must; or he must be given a microgrammar which has been designed in such a way as to have boundaries which the user very seldom oversteps.

---

11   No matter how small, of course, if the microgrammar contains recursive loops the set of specified sentences, being infinite, cannot be memorized as such.

12   Many computers have been programmed so that their users can type in such instructions as 'PRINT' or 'COMPUTE' or whatever. Using English words in this way hardly constitutes using a microgrammar, - if that term is to mean anything - any more than a Dubuque housewife buying a 'chaise longue' is using French.

Of these two alternatives, clearly the second is preferable, for the first would be at best a painful process, and in effect an interminable one[13].

To point up the difference between microgrammars and ALGOL, we may note that the Frenchman of our example has much less trouble staying within the bounds of the portion of English he shares with the American than the American does. That is, the Frenchman too will tend to extrapolate into areas beyond the boundaries of that portion, both on the basis of what English he knows and on the basis of French, insofar as he feels that his fragmentary English "corresponds" to French. And the Frenchman will, from time to time, by accident hit on a well-formed English sentence. But obviously his situation is quite different from the American's: he is extrapolating into a foreign language, on whatever basis, and is presumably almost always aware of the fact that in effect he is "creating" new and unfamiliar expressions[14]; while the American, for his part, is comparing two sentences from a language he is thoroughly familiar with, and deciding that the second is enough 'like' the already-accepted first to warrant trying it. The Frenchman I think will be less likely to extrapolate, he will be more self-conscious about it; it should then be far easier for him to stay within the boundaries of the English portion he shares with the American. His situation is not unlike

---

13   Actually a third alternative might be thought of, one suggested by our example of the American and the Frenchman. That is, just as the Frenchman does not really have to understand a sentence in its entirety to get the gist of what's being said, so too a microgrammar might be able to analyze whatever it is equipped to, ignoring the rest or dealing with the remainder in accordance with the portion previously parsed. (In present practice if a sentence cannot be wholly digested, it's disgorged.) This alternative is taken up again in a later paragraph, q.v.

14   More exactly, 'new and partly unfamiliar expressions' --- they must be partly familiar in order for him to 'extrapolate' them.

that of someone who knows ALGOL: the ALGOL-user, too, must occasion-
ally 'extrapolate' new pseudo-ALGOL expressions, on the basis of what
he knows about ALGOL; but in general he is less likely to do this
than the American is to extrapolate as 'within the shared portion'
sentences he knows to be well-formed English utterances[15].

In sum, the thesis presented here is that it is easier to
avoid extrapolating 'new' expressions than 'familiar' ones[16].
If this is true, it will be easier to devise an ALGOL for careless
(i.e. human) users than to devise a microgrammar for such users:
fewer mistakes must be anticipated.

Which returns us to our main point: if a useful microgrammar is
to be provided, one must be devised which allows in advance for the
extrapolation it will inevitably provoke; a microgrammar which relies

15   This point should not be overemphasized, however; ALGOL-programmers
do after all make mistakes (not all of them extrapolative); the lesser
likelihood of their doing so is of no great service to them.

16   I should note that a point similar to the one I have made with
my hypothetical American and 'shared portion' of English, has been
made by F. W. Alt using for his example an English-speaker trying
to stay within Basic English. Thus, on page 137 of his Electronic
Digital Computers ((1)), Alt speaks of "Basic English, that artificial
language in which only a small number of English words are used:
it is easy to learn for someone who does not known (sic) English,
but it is quite difficult for an English-speaking person to learn
to avoid the forbidden words." As opposed to Alt's emphasis on
"forbidden words", however, I have tried to stress the element of
"forbidden structures". It is far easier to expand a microgrammar
so as to specify more "words" than it is to expand it so as to well-
specify more sentence-types. I should also point out that, although
both Alt's thesis (regarding words) and mine (regarding sentences)
are highly plausible, I know of no experimental verification of either,
nor does Alt cite any. Such a test should be far from difficult to
devise, however.

on extrapolation as the only means of mastering it (since its rules cannot be learned), must certainly provide for the liabilities of the extrapolative process[17].

To state this problem is to state in brief the problem of making microgrammars which are already linguistically-adequate, useful for artificial-intelligence systems.

What are the properties of an English portion which users can stay within -- of "habitable" English fragments? One way of finding out is to discover what 'extrapolative' mistakes are likeliest, for this knowledge might be used to incorporate into a microgrammar the areas into which 'extrapolation' most often penetrates, thus making the expanded 'anticipatory' microgrammar habitable[18].

Without making any claim that there is an exact classification of the kinds of 'extrapolation' to be expected, I think it might still be profitable to distinguish between 'syntactic' extrapolation (often paraphrastic)[19] and 'lexical' extrapolation; and, under each of these headings, between 'close' and 'loose' extrapolation.

---

17   The contradiction here is only apparent. If a microgrammar, no matter how much extrapolation it included, endlessly fomented further extrapolation to the same degree, then our work would never be done, and the 'extrapolation-including' microgrammar would be an impossibility. From a given lexicon and a given set of rules, however, not all English sentences are likely to be 'extrapolated'; very few, comparatively, are likely to be extrapolated often. On this assumption, at least, rests our hope of building a microgrammar which, 'providing for' high-frequency extrapolations, facilitates the user's staying within the bounds of its microlanguage.

18   It may also be possible to accommodate many types of non-extrapolative mistakes -- clerical errors and so on -- but this possibility will not be further discussed here.

19   But not always ('Here are some neurons' ≠ 'Here is a neuron'); and often a 'degree of paraphrase' must be admitted, informally, as in the examples which follow.

Examples follow:

TYPE OF 'MISTAKE'	EXTRAPOLATING FROM:	TO:
1. Syntactic, close	Is a neuron here?	Is there a neuron here?
2. Syntactic, loose	Is a neuron here?	a. Are there any neurons here? b. Do there happen to be any neurons here? c. Can you find any neurons here? d. Do you see anything in the way of a neuron here?
3. Lexical, close	Is a neuron here?	Is an astrocyte here?
4. Lexical, loose	Is a neuron here?	a. Is a cell here? b. Is a larynx here? c. Is a gizzard here? d. Is a zeppelin here?

I would rather not devote very much attention to these ad-hoc distinctions, but a brief exposition might be in order. All of the cases of 'syntactic' extrapolation are pariphrastic, at least to a very high degree:[20] the user, on the basis of what the microgrammar

---

20 That is, their extensions largely overlap.

has already accepted, analogizes as equally acceptable a similar

sentence which to him means more-or-less the same thing, under

identical contextual conditions.[21] In the case of 'close' syntactic

extrapolation, the paraphrase is first of all seemingly exact, and

is also extremely close in structure to the original; in fact, I

see only a stylistic difference between them.[22] I think these first

two sentences are also closer together semantically than are the

first and any of the 'loosely'-extrapolated sentences; but since

this point is unimportant I have not tried to devise any tests to

subject this claim to scrutiny. The 'loose' syntactic extrapola-

tions are ranked by increasing 'laxity,' or difference in form from

---

21   Many of the 'extrapolations' characteristic of the situations
    here under view must be cases of analogy: for example, where
    sentences 'a,' 'b,' and 'c' are in the accepted portion, and
    'b' is similar to 'a' in some (naive) way, then an 'analogical
    extrapolation' might produce some 'd,' presumed also to be
    within the accepted portion, on the analogy "$\underline{a}$ is to $\underline{b}$ as $\underline{c}$ is
    to $\underline{d}$." Or, between two forms only, the analogy might be taken
    as "$\underline{a}$ is to the accepted portion as (the similar) $\underline{b}$ is." I
    want to avoid taking this point much further here, both
    because the term "extrapolation" is less specific than "analogy,"
    and is thus more suitable to the present status of this
    research; and also because I want to avoid the implication that
    the 'extrapolative' process here treated is to be identified
    with the process of 'analogical change' as studied by
    diachronists, e.g., in ((6)). There are, to be sure, some
    points of apparent similarity between the two processes, but
    it would be premature to attempt to state their kinship.

    I have also wanted to avoid bearing down too hard on the
    'paraphrastic' elements here alluded to; for a more serious
    study of paraphrase, see e.g. ((5)).

22   On the other hand, it might be contended that sentences like
    "Is a neuron here?" are so deviant as to be non-English: I
    have heard an eminent linguist defend this view, which may well
    be valid for his idiolect.

the original. (2a) is rather close in form, and may be identical in meaning. (2b) is rather distant in form, and seems to involve additional meaning elements; (2c) is at least as distant in form, and has added mention of the addressee ("you"); and (2d) is very far in form and is couched in extremely informal style.

The cases of 'lexical extrapolation' involve substitution of one word (=lexical item) for another.[23] In the example of 'close' lexical extrapolation, 'neuron' is replaced by 'astrocyte:' both are names of cells found in e.g. the human brain. In the 'loose' lexical extrapolations the substitutions can be intuitively judged to result in greater semantic distance. (4a) substitutes 'cell,' which is a generic term for the class one of whose members is 'neuron;' (4b) substitutes the name of a quite different body-part, though still preserving medical terminology; (4c) substitutes, again, the name of a body-part, but in the vernacular; and (4d) substitutes a noun chosen practically at random.[24]

---

23 In generative grammars like PLACEBO IV, the difference between 'syntactic' rules and 'lexical' ones is a difference only between locations in the instantiative path: 'lexical' rules are last to be actuated in generation, first in analysis (parsing). PLACEBO IV does not distinguish the lexical level by pausing prior to reaching that level for a 'change-of-gears' into a context-recognizing model; in fact only in the most trivial sense can PLACEBO IV be said to have morphographemicizing rules at all.

Even in grammars which do pause for context-recognition, however, the difference between 'syntax' and 'lexis' needn't be very marked, in the sense of there being artificial 'levels' imposed on the grammar.

24 Or, more accurately, at random from the set of inanimate count-nouns.

From the above discussion I think it is obvious that if no constraints whatever were placed on the use of the microgrammar, and if the user were at all inclined to stray into the areas delimited above: that there would be little hope of containing him within anything short of an English qcg. However, the user will not normally be so reckless. Factors tending to restrain his discourse are:

1. The fact that he is seated at a computer console;

2. His being there to discuss a given field of knowledge;

3. His having been warned that his addressee, like our hypothetical Frenchman, 'understands a little English.'

The first and third factors will operate together with the primary effect of limiting the syntactic rules the user calls into play. In a formal situation he will be less casual than if he were addressing an idler in the Courthouse Square; his discourse would already be somewhat restrained if he were addressing a human colleague. Also, seated at the keyboard, and hopefully being less than a practiced typist, he will tend to avoid lengthy circumlocutions and convolute syntax. The brevity of his expressions will be increased still more, perhaps, by his knowledge that the computer he is addressing is an expensive interlocutor.

The first and third factors, in conjunction with the second, should serve to restrict the user's vocabulary. The extent to which the second factor plays a role in this regard may well depend, of course, on which field of knowledge is at issue; but it would be

mere guesswork to attempt at this time to give even a crude ranking
of professional dialects with respect to how much English each calls
on.[25]

All of the factors cited above can and should be exploited by
the microgrammarian as facilitating his task. In particular, these
factors should serve to reduce to the vanishing point some of the
'loose' extrapolations listed above: (2b), (2c), and (2d) should
no longer be problems, and certainly (4c) and (4d) should not be.
(4b) may constitute a problem in inverse proportion to the degree of
specialization of the user, or of the subject he is concerned with;
a general practitioner might have more cause to use the word
'larynx' than a neuropathologist, for example. Left as principal
problems, then, are four kinds of extrapolation: 'close syntactic,'
least-lax 'loose syntactic;' 'close lexical,' and least-lax
'loose lexical.' It is probable that all of these will plague the
microgrammarian: and, barring steps to overcome them, the user of
the system.

------

25   English may be regarded as being composed of, or manifested in,
      many different dialects (each in turn composed of, or manifested
      in, many different idiolects). There are a number of ways of
      subdividing English into its dialects; these ways are mutually
      conflicting. Chiefly, one may delimit dialects of area, of
      class, and of profession. The first of these is under extensive
      investigation the world over, in the form of the many 'Linguistic
      Atlas' projects under way. The second has received somewhat
      less attention. The third has received still less; though
      studies are often published (as in the journal American Speech)
      of professional and cant terms, I know of no study of the
      varieties of syntax distinguished by various professional
      dialects.

The most obvious way of overcoming them, and the dullest and least practical, is to station a linguist at the user's shoulder, instructing him to expand the grammar whenever its limits are exceeded. Next most obvious, and extremely challenging, is to provide a mechanical device to perform this linguistic service. This device would analyze as much of the sentence as it could, then integrate the remainder into the grammar (instead of merely suppressing it) and parse the sentence in its entirety.[26] The most elegant and satisfying way of overcoming the difficulties, and also one of some linguistic or at least psycholinguistic interest, is to anticipate them in the microgrammar itself. To continue the example treated above, a microgrammar which specifies "Is a neuron here?" should also specify "Is there a neuron here?", "Are there any neurons here?", "Is an astrocyte here?", and "Is a cell here?". And therefore, of course, also "Are there any neurons here?", "Is there an astrocyte here?", and so on.[27]

---

[26] The Klein-Simmons program described in ((7)) can be looked on as essentially a device of this kind. Its computations (from sentential context) are quite limited in scope, however (without questioning the efficacy claimed for them); and a much more powerful program would be necessary to carry out the operations necessary to allow the computer to surrogate the 'over-the-shoulder linguist.'

[27] Before we close this subject, perhaps it should be remarked that in practice we could expect syntactic and lexical violations to occur simultaneously; a user might well extrapolate from "Is a neuron here?" to "Are there any astrocytes here?", for example.

It would seem not unreasonable to group all of these factors together as comprising a single prerequisite of 'habitable' microgrammars: a prerequisite we might call 'extrapolative symmetry.'[28] I do not propose to define this term very exactly here, but its meaning should be intuitively obvious from the preceding discussion: a microgrammar which has 'extrapolative symmetry' will contain no extrapolative gaps, such as the omission of "Is there a neuron here?" in an algorithm including that sentence's fellows. Naturally no microgrammar will possess complete extrapolative symmetry---for in any case the term has been left indefinite enough so that this absolute quality might be hard to recognize---; but the measure to which it approaches this ideal will be, I suggest, the measure of its utility for human users.[29]

---

28   Symmetry as a general property has been much discussed as among
      the desiderata of linguistic analysis. Harris ((4)) has treated
      this desideratum with regard to all levels of analysis; more
      often it is treated with regard solely to phonological analysis,
      where the doctrine has recently become somewhat controversial,
      as see e.g. ((8)) and ((10)).

29   If the process of 'extrapolation' had been defined with any
      precision we could discover whether or not a microgrammar (or a
      natural language) could have complete extrapolative symmetry:
      whether, that is, eventually a point may be reached where new
      extrapolations produce only sentences already in the language
      (or microlanguage). I know of no evidence to support either
      this conclusion or its converse.

      In any case it should be pointed out that PLACEBO IV obviously
      lacks complete extrapolative symmetry; so also will its
      successors. I hope only that it will ultimately be possible
      to produce a microgrammar which approaches 'complete
      extrapolative symmetry' closely enough to significantly reduce
      the amount of trespassing that takes place during use---and
      also the number of injunctions that must be kept in mind by
      users.

I have argued that if a given microgrammar has 'extrapolative symmetry' in large measure, and if the user of that microgrammar is mindful of the general constraints placed on him: that he will be able to make new sentences without often transgressing the limits of the microlanguage. In this respect his behavior, and his freedom, will resemble those of the speaker of English, who also makes new sentences without often transgressing the limits of the language. It may be tempting to draw the inference that the way in which English-speakers form new English sentences is somewhat akin to the way in which microgrammar-users form new sentences in the microlanguage: that is, that both use an extrapolatory process.[30] Such a conjecture however has no bearing on the point at hand, which is that regardless of what processes are used, the English-speaker

---

30 On the other hand there are marked dissimilarities between the behavior of English-speakers and the (expected) behavior of microlanguage-users. Not least of these is the fact that microlanguage-users necessarily create their sentences in an artificial situation, without ever quite forgetting that artificiality; whereas English-speakers are scarcely aware that they are 'using a grammar' to produce their sentences. For this reason, and for others, I would like to stop short of making the above-cited conjecture.

must be enabled to carry over into his use of a microlanguage some of the same freedom he has enjoyed in his use of English.[31]

In insuring that his microgrammar have 'extrapolative symmetry,' the linguist's first task will, obviously, be to make certain that for each sentence structure in the microgrammar, all of the <u>common</u> paraphrasing structures are also included in the microgrammar. Thus, if the microgrammar specified sentences such as "A neuron is to the left of an astrocyte," it should definitely include sentences of the form "There is a neuron to the left of an astrocyte," "An astrocyte is to the right of a neuron," and so on. (With respect to increasing the utility of the microgrammar, it does not matter that we have only a rough idea of what 'paraphrases' are; as native speakers of English our 'rough idea' is likely to correspond to the user's, and in any case the microgrammar does not <u>label</u> (for the most part) any sentences as paraphrastic, so that if the linguist merely tries to include all common <u>apparently-paraphrastic</u> sentence-types,

---

31   In the unconstrained 'everyday' situation a speaker is almost totally unaware of the restrictions placed by English on what he wants to say. (Whether this is because English is beautifully adapted to expressing what he wants to say, or because what he wants to say is conditioned by English, is a moot question.) But what is more immediate to our present concerns is the fact that under the same circumstances that speaker is almost totally unaware of the restrictions placed by English on <u>how</u> he will express what he wants to say. It is surely an essential part of our linguistic habits that we are not constantly fumbling for the right grammatical rule, in our effort to express something or to correct a malformed expression. This is to say, without putting any weight on a gratuitous assumption about the English-speaker's 'extrapolative' activities, that if such speakers <u>did</u> make all new sentences on an 'extrapolative' basis, then English would have to have a very large measure of what I have called 'extrapolative symmetry.'

he can hardly go wrong.) But this will not be enough, if our idea

of 'paraphrastic' is 'truth-preserving,' for as I have gone to such

lengths to point out, the microgrammar-user may be expected to

extrapolate from "A neuron is next to..." to "Neurons are next to...".

Thus, a secondary requirement must be met: that common nonpara-

phrastic extrapolations (or paraphrases which are only <u>locally</u>

(in the context at hand) truth-preserving) be included in the

microgrammar. This requirement may boil down to a rule that certain

classes of <u>transformational</u> <u>rules</u> must be taken into account when

building a microgrammar, in that certain resulting <u>transforms</u> must

be included. Should this be true the process of building a

microgrammar of 'extrapolative symmetry' will have been made less

difficult, because more understood; but it is still too early to

venture a prediction on how likely the microgrammarian is to have

this good fortune.[32]

---

32 Obviously, the more transformational relations a given transform
    grammar recognizes, the more likely that grammar is to provide
    the microgrammarian with information on what 'extrapolations'
    his product ought to include. A transform grammar which ignores
    the relation between "He kicked her" and "He gave her a kick,"
    for example, would not suggest that a microgrammar including
    the first should include the second. Generally speaking
    transform grammarians of the Harrisian school feel more
    constrained to treat <u>all</u> intersentential relations of a
    transformational nature (though this insistence is due as much
    to predilection as to doctrine); grammars from this school, then,
    may be of greater usefulness to microgrammarians. This point
    may also be taken as an argument against using any but the
    transformational model for microgrammars. The argument is
    cogent (as are others on this point); but for the moment it is
    necessary to make do with what we have.

2. Conclusion and Prospect.

I hope to have shown that whereas against qcg microgrammars have at least the immediate advantage that they can be made available now, against e.g. ALGOL microgrammars have the permanent advantage that they will allow their users to carry over some of their general linguistic habits into their communication with computers. I have suggested however that these same linguistic habits will constitute a serious stumbling-block to the utility of microgrammar unless these algorithms can offer, in extrapolative freedom, an analog to the latitude which qcg seem to have. I have indicated how this freedom may be possible of attainment, through its being restricted in scope by the parameters of the using situation, and through its being provided, when within these parameters, by a Klein-Simmons type of auto-grammarizer and/or by the incorporation of extrapolative symmetry. I have sketched some guidelines for suggested future research.

Some of this research is already underway. PLACEBO IV is gradually being expanded, one of the chief objectives of this work being provision of a microgrammar of high extrapolative symmetry.[33]
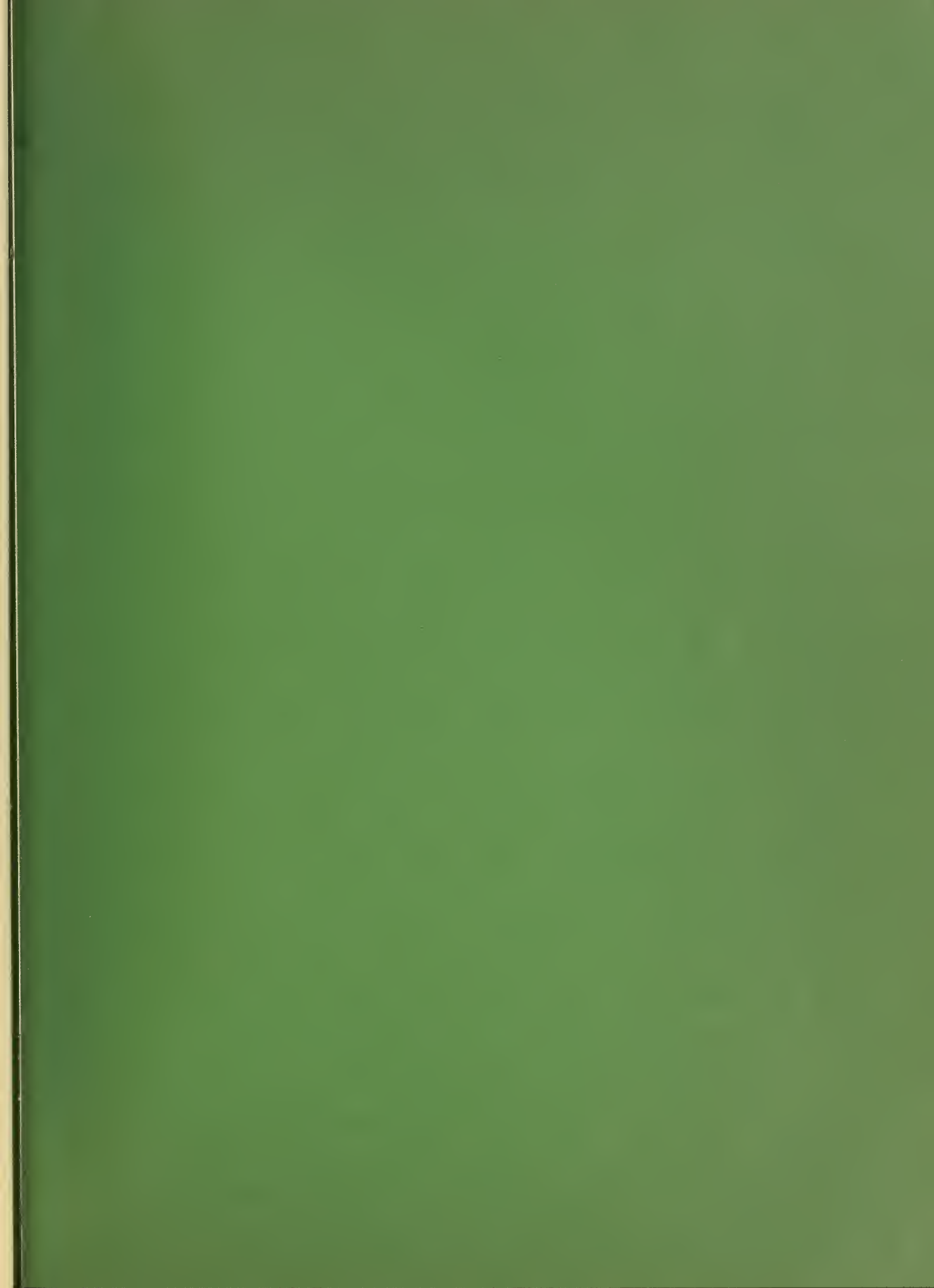
---

33 It may well be that the end result of successively approximating a given microgrammar to 'maximal extrapolative symmetry' is necessarily a qcg of English, or at best an English qcg less some of its terminals (e.g. 'zeppelin'). If true, this is irrelevant, for the microgrammarian is not out to supply full extrapolative symmetry, only enough to make the microgrammar 'habitable.' Knowing in advance how much 'symmetry' will be necessary for this end would be extremely difficult, but such knowledge will not be necessary; the linguist can test the micro-grammar against user-habitability at many points during the time of development; and he can, at some more-or-less arbitrary point, decide (with the user's compliance) that the microgrammar has become habitable enough to warrant cutting off further develop-ment.

Experiments with informants are to be conducted to ascertain how close to reaching habitability each stage of the microgrammar has come, and to learn where informants overstep the microgrammar's bounds.

It may turn out that my hope that microgrammars can be made habitable is unfounded. In light of this possibility, I think it must be said that enough is now known about microgrammars and about the prerequisites for their utility, to predict that if they cannot be designed in such a way as to permit transference to their use of general linguistic habits, with little penalty for such transference, they will necessarily be relegated to the status of curiosities.

REFERENCES

((1)) Franz W. Alt, <u>Electronic Digital Computers</u>, pp. x, 336.
New York and London (1958).

((2)) Noam Chomsky, <u>Syntactic Structures</u>, pp. 116.
<u>Janua Linguarum, Studia Memoriae Nicolai Van Wijk Dedicata</u>, IV
's Gravenhage (1957).

((3)) Zellig S. Harris, <u>The Elementary Transformations</u>, pp. 110.
<u>University of Pennsylvania Transformations and Discourse
Analysis Papers</u>, #54. ('Pre-publication copy').
Philadelphia (1964).

((4)) Zellig S. Harris, <u>Methods in Structural Linguistics</u>, 1951;
Fourth Impression pp. xvi, 384.
Chicago (1960).

((5)) Henry Hiz, <u>The Role of Paraphrase in Grammar</u>, pp. 12.
<u>University of Pennsylvania Transformations and Discourse
Analysis Papers</u>, #53.
Philadelphia (1964).

((6)) Henry M. Hoenigswald, <u>Language Change and Linguistic
Reconstruction</u>, pp. viii, 168.
Chicago (1960).

((7)) Sheldon Klein & Robert F. Simmons, "A Computational Approach
to Grammatical Coding of English Words."
J.ACM 10.334-347 (1963).

((8)) Kemp Malone, "On Symmetry in Phonemic Analysis."
<u>Lg</u> 38.142-146 (1962).

((9)) Jane J. Robinson, <u>Preliminary Codes and Rules for the
Automatic Parsing of English</u>, pp. ix, 140.
The RAND Corporation, Memorandum RM-3339-PR.
Santa Monica (1962).

((10)) W. Freeman Twaddell, "Uses and Abuses of Symmetry."
<u>Studies in American English: Second Texas Conference on
Problems of Linguistic Analysis in English</u>; cited article
pp. 129-137, discussion pp. 137-150 and 162-163.
Austin, Texas (1962).

((11)) William C. Watt, <u>General Properties of Microgrammars</u>.
(forthcoming).

((12)) William C. Watt, <u>PLACEBO IV: Rules, Concordance, Sample
Computer Generation</u>.
NBS Technical Note 255
Washington (1965).