NATURAL LANGUAGES AND THE CHOMSKY HIERARCHY

András Kornai
Institute of Linguistics
Hungarian Academy of Sciences
Budapest, P.O.Box 19, H-1250 Hungary

Abstract

The central claim of the paper is that NL stringsets are regular. Three independent arguments are offered in favor of this position: one based on parsimony considerations, one employing the McCullogh-Pitts (1943) model of neurons, and a purely linguistic one. It is possible to derive explicit upper bounds for the number of (live) states in NL acceptors: the results show that finite state NL parsers can be implemented on present-day computers. The position of NL stringsets within the regular family is also investigated: it is proved that NLs are counter-free, but not locally testable.

0 Introduction

The question whether the grammatical sentences of natural languages form regular (Type 3), context free (Type 2), context sensitive (Type 1), or recursively enumerable (Type 0) sets has been subject to much discussion ever since it was posed by Chomsky in his seminal 1956 paper. However, there seems to be little agreement among the linguists concerned with the 'geographic' position of natural languages (NLs): for instance Reich (1969) claims NLs to be finite-state (Type 3), while Matthews (1979) argues that they are not even recursively enumerable.

Pullum and Gazdar (1982) have demonstrated that the standard *linguistic* arguments against the context-freeness of natural languages are fallacious – they did not consider, however, the *metalinguistic* argument offered by Matthews. In Section 1 of this paper I will briefly outline and challenge this argument, and in Section 2 I will argue in favor of Reich's position. The claim that NLs are Type 3 has several implications for linguistic (meta)theory: these will be discussed in Section 3.

The paper presupposes some familiarity with the basic notions and notations of formal language theory: when no specific reference is given, the reader will find a proof both in Salomaa (1973) and Harrison (1978)

1 Natural languages as formal languages

The *extensional* view of natural languages, i.e. the identification of NLs with the set of their grammatical strings (sentences) is sometimes regarded an idea characteristic of generative linguistics. Since it was Chomsky (1957:18) who first made this view explicit, this is not wholly unjust: yet it is quite clear that the same idea was implicit in much of the work of the structuralist period.[1] In fact, the 'discovery procedures' developed by the structuralists in order to arrive at a concise description (grammar) of a NL from a set of utterances (corpus) were without exception based on the assumption that native speakers of the language are capable of judging the grammaticality of utterances presented to them. Although these procedures are, by and large, practical (empirical) and mechanical (algorithmic), their presentation already involved a certain amount of idealization.

For instance, it is obvious that native speakers themselves utter ungrammatical sentences from time to time, and it is also clear that they can understand (parse) sentences that are not strictly 'grammatical'. Nevertheless, these methods work quite well in the actual practice of NL description, and the structuralist methodology has often been compared to that of chemistry, physics, and other natural sciences.[2] Matthews (1979) has casted

[1] See e.g. Def 4 in Bloomfield 1926, or Harris 1946 ch 1.0

[2] See e.g. Carroll 1953, Levi-Strauss 1958 ch. 2

doubts on the fundamental assumption of these procedures: he claims that native speakers are in fact *unable* to judge the grammaticality of material presented to them. The relevant part of his argumentation is reproduced below:

"Consider (1) and (2). If native speakers instantiate

(1) the canoe floated down the river sank

(2) the editor authors the newspaper hired liked laughed

an effective procedure in their classification of sentences, then presumably the classification of (1) and (2) should not depend on their position in a list of test sentences that also includes sentences similar to (3) and (4).

(3) the man (that was) thrown down the stairs died

(4) the editor (whom) the authors the newspaper hired liked laughed

but in fact it does. (1) and (2) will typically be classified as ungrammatical if they precede sentences similar to (3) and (4), but grammatical if they follow them. Such cases are quite common." (p 212)

Moreover, "there is considerable empirical evidence to suggest that native speakers employ a battery of heuristic strategies when parsing and classifying sentences. Their reliance on such strategies does not preclude their having available to them an effective procedure for deciding membership in their language; however, in the absence of empirical evidence for such a procedure, we are certainly not required to postulate its existence." (p 213)

From this, Matthews concludes that Putnam (1961) was not justified in appealing to Church's thesis in order to show that NLs are recursive: for if native speakers have no effective procedure for deciding membership in the set of grammatical sentences, then there is no guarantee that such procedure exists. But is it really the case that the "battery of heuristic procedures" employed by native speakers falls outside the scope of Church's thesis? Well- confirmed natural laws[3] are usually taken to be universally valid – it is unclear why should Church's thesis be an exception to this, and Matthews offers no evidence to corroborate his views on this point.

Putnam's original argument derives its strength from Church's thesis. If NLs are *not* Type 0, then the heuristic strategies of native speakers will be instances of precisely that sort of procedures that Church's thesis predicts *not* to exist: on the one hand, they are 'intuitively effective', and on the other hand, they are not Turing computable.

The phenomenon observed by Matthews, namely that native speakers can be coaxed into accepting (or rejecting) sentences has little to do with the recursiveness of the battery of heuristics they employ: rather, it calls the extensional view of language in question. The problem is a methodological one: if NLs are defined to be sets of grammatical sentences, how can one test potential elements for membership? This problem becomes particularly acute in borderline cases (such as (1-4) above), and for the linguist who wants to check the predictions of his grammar it matters but little that such dubious sentences are (statistically) infrequent.

The easiest way to solve this problem is to give up the assumption that grammaticality is a yes/no question: 'degrees of grammaticalness' can be introduced (see e.g. Chomsky 1961) and NLs can be treated as graded (or even fuzzy) sets. This approach, however, can only be applied in the study of idiolects (languages of individual speakers), because there is no way to arrive at a graded set that will reflect the sum of individual opinions in a reasonably faithful manner.

Suppose, for instance, that we have three speakers, X, Y, and Z, and each of them classifies the sentences a, b, and c consistently (that is, if he prefers a to b and b to c, then he prefers a to c, etc.). Now, if for speaker X $a > b > c$, for Y $b > c > a$, and for Z $c > a > b$, then the majority prefers a to b, b to c, and c to a; in other words, the 'general opinion' is inconsistent (non-transitive). The "possibility theorem" of Arrow (1950) makes it clear that the example is typical: under very general conditions, there is simply no way to aggregate graded sets in such a manner that the (partial) orderings imposed by the individual gradations are preserved. Therefore, the 'degrees of grammaticalness' approach must be relegated to the study of idiolects in any case – most linguists, however, reject it entirely (see Newmeyer 1980 ch. 5.5.2, 5.7.1).

Yes/no grammaticality judgments, on the other hand, show remarkably little variation from speaker to speaker in any given speech community, and it is this *intrasubjective testability*

[3] For independent motivation of Church's thesis, see e.g. Rogers 1967: ch. 1.7

(cf. Itkonen 1981) that justifies the empirical study of 'dialects' and even 'languages'. But if Matthews is right, and native speakers are unable to "classify any sentence over the vocabulary of their language *consistently* as either grammatical or ungrammatical" (p 211), then intrasubjective testability will be impossible to achieve. The question is: what makes the native speaker inconsistent? In Matthews' example, there can be little doubt that the cause of the inconsistency is the *test situation:* the speaker's linguistic intuition is not the same before and after reading sentences (3-4).

This source of inconsistency can be eliminated fairly easily: if the sentences are presented in a random manner (preferably, with "filler" sentences among them), then no "cues provided by the context of classification" (p 213) will be present. Naturally, linguistically relevant experiments will have to control many other factors (see e.g. Greenbaum and Quirk 1970), but as we shall see, there is no need to discuss these individually.

From the point of intrasubjective testability, it can be safely said that well-designed experiments usually provide highly consistent data (even in the case of borderline sentences), and the extensional view of NLs can be maintained on an empirical basis as well. The actual sets designated as NLs will, at least to a certain extent, depend on the choice of experimental technique, but any *fixed* experimental method can be thought of as an algorithm for deciding questions of membership *with the aid of a human oracle.*[4]

Since the existing experimental methods can be replaced by (interactive) computer-programs, the question boils down to this: is a Turing machine with a human oracle more powerful than one with a Turing machine oracle? By Church's thesis, the answer is negative, and as Turing machines with recursive oracles are no more powerful than Turing machines without oracle (see e.g. Rogers 1967 ch. 9.4), NLs must be recursive.

Notice, that this line of reasoning is independent of the particular choice of experimental technique, or what is the same, of the precise definition of NLs. This is a consequence of the fact that the experimental methods used in empirical sciences (including linguistics) hardly merit this name unless they are well-defined and 'mechanical' to such an extent that their algorithmization poses no real problems. (For instance, the procedure outlined above does not make crucial reference to random sequences: the 'randomization' of test-sentences can be carried out with the aid of pseudorandom sequences generated by Turing machine.) This is not to say that introspective evidence or intuition plays no role in linguistics (or in general, in the development of science) – but questions concerning the position of natural languages in the Chomsky hierarchy can hardly be meaningful unless we have some definition of NLs (i.e. some experimental method to test membership) to work with.

[4] For the definition of oracles, see e.g. Rogers 1967 ch. 9

2 The regularity of natural languages

Finite state NL models were first developed by Hockett (1955). Although Chomsky (1957 ch 3.1) attempted to demonstrate the inadequacy of such models, several linguists[5] advocated their use, and the stratificational school of linguistics (Lamb 1966) persists in employing a formalism which is, in essence, equivalent to finite automata (cf. Table 1 of Borgida 1983).

As Reich (1969) has pointed out, Chomsky's demonstration is based on the assumption that NLs are self-embedding *to an arbitrary degree.* This means, that the sentences (1-2) and (5-6) must be equally grammatical:

(5) the boss editor authors the newspaper hired liked hates laughed

(6) the committee boss editor authors the newspaper hired liked hates chairs agreed

The experiments (Miller and Isard 1964, Marks 1968), however, do not support this conclusion: native speakers of English react to (5-6) and (7-8) the same way[6]

(7) the boss editor authors the newspaper hired liked hates laughed cursed

(8) the secretary committee boss editor authors the newspaper hired liked hates chairs agreed

Since (7-8) are ungrammatical in any grammar of English, Chomsky's original demonstration is far from convincing, and the question whether NLs

[5] Especially the ones working with computers. See e.g. Marcus 1964, Church 1980

[6] Chomsky (1963) regards (5-6) grammatical (but unacceptable) and (7-8) ungrammatical: for the methodological implications of this position see Greene (1972).

are Type 3 is still open.

In fact, the only way to show that NLs are not *finite* is to exhibit some infinite sequence of grammatical sentences: fortunately, the pattern exemplified in (1-6) is not necessary for this. Coordinated constructions as in

(9) I have seen Tom

(10) I have seen Tom and Dick

(11) I have seen Tom, Dick and Harry

can be as long as we wish: the grammaticality of such sentences is independent of the number of conjuncts. Similar (right- and left-recursive) patterns can be found in any NL, but all of these can be described by regular expressions. Therefore, if grammars do not have to account for iterated self-embeddings, the principle of scientific parsimony will point to the *minimal* language family accommodating every possible finite NL corpus and their regular extensions. From this perspective, the Type-3 family is more than sufficient: since it contains every finite language and is closed under regular operations, it provides a generous upper bound for the family of NLs.

A more direct argument can be based on the biological make-up of the human brain: as individual neurons can be modeled by finite automata (McCulloch -Pitts 1943), and a finite three-dimensional array of such automata can be substituted by one finite automaton (see Kleene 1956), NLs must be regular. Although finite state models of NLs usually do not claim "neurological reality" (see ch 3.2 of Sullivan 1980), the above reasoning gives us an *upper bound* on the complexity of finite automata necessary to describe NLs: since the relevant part of the brain contains no more than 10^8 cells, and one cell has cca. $10^2 - 10^3$ states, non-deterministic automata with 10^{10^9} states will be sufficient.

Since the neurological organization of the human brain is unlikely to parallel the actual organization of the (internalized) grammar of native speakers, it is not surprising that the application of linguistic methods gives a much sharper upper bound: as we shall see, finite deterministic NL acceptors need not have more than 10^{16} states. This estimation can be derived from the investigation of the syntactic monoids defined by NLs. (For the definition of syntactic monoids, see McNaughton–Papert 1968, and for a systematic exposition, see ch 3.10 of Eilenberg 1974.)

Elements of the syntactic monoid correspond to the *distributional classes* of structuralist linguistics: two strings will belong to the same class if and only if they have the same distribution, i.e. iff they can be substituted for each other in any sentence of the language in question. The distributional classes formed by strings of length one will be the elements of the *terminal* alphabet; but it should be kept in mind that these function as preterminals inasmuch as each of them stands for a (not necessarily finite) class of elements. In a morpheme-based approach, terminals are called *morpheme classes:* these can be set up by the procedure outlined in ch 15 of Harris (1951).

In a word-based approach, that is, if we take words to be the ultimate syntactic constituents, the terminals will be called *lexical (sub)categories:* in either case,[7] the number of terminals is clearly finite. There are no more than 20 lexical categories; and in any given NL there are less than 300 morpheme classes. However, fully formed words with different inflexional affixes will belong to different distributional classes, and if we want to maintain the regularity of lexical insertion, lexical entries (e.g. verbs or verb stems) with different subcategorization frames will fall in different subclasses. Traditional accounts of lexical (sub)categorization also allow for overlapping classes (in cases of homonimy). For the sake of simplicity, I will take the *Boolean atoms* of such systems as basic: this way, elements like 'run' or 'divorce' will be neither nouns nor verbs but will be listed under a separate category for 'noun-verbs'.

But even if we take all these factors into account, it can be safely said that the number of morpheme classes does not exceed 10^3 and the number of lexical subcategories does not exceed 10^4 in any given NL. In other words, it is possible to select for any given NL a 'core vocabulary' (or morpheme list) of $10^4(10^3)$ elements in such a manner that every word (morpheme) not appearing in the list will be distributionally equivalent to one already on it. This means that the number of states that can be reached in one step from a given state of a finite state NL acceptor cannot exceed 10^4 and conversely, any given state can be reached from at most 10^4 states in one step.

The states of finite automata are in one-

[7] Present-day syntacticians seem to favor the latter approach: for discussion see Robins (1959), Chomsky (1970), Lieber (1981).

to-one correspondence with the classes of right-distribution: two strings over the terminal vocabulary will take the (minimal) automaton to the same state iff they can be substituted for each other in every right-side environment. As a special case, it should be mentioned that those strings that do not appear as initial parts of grammatical sentences will give only one state in the automaton: these, therefore, can be disregarded. The remaining strings (i.e. the ones that can be finished grammatically) have a property peculiar to NLs: they can always be finished with *at most* four words (or twelve morphemes).[8] This means that the final state[9] of NL acceptors can be reached from every live state in at most four (twelve) steps. Therefore, the number of live states is at most $10^{16}(10^{36})$.

3 Consequences

It should be emphasized that the above estimation is still very generous: a systematic study of sentence endings is highly unlikely to reveal more than 10^5 different patterns in any given NL, and the proper order of magnitude seems to be 10^4. If the automaton has to account for the morphology of the language as well, $10^6 - 10^7$ states will be necessary – this is, perhaps, outside the capabilities of present-day computers. In any case, finite automata can be implemented on *any* theoretical (or actual) model of serial computation like Turing machines, random access machines, etc. to accept languages in *linear* time.

Although native speakers understand grammatical sentences in *real* time, their performance as NL acceptor is somewhat hindered by the fact that the heuristic algorithm they use is not adopted to ungrammatical strings: usually they spend some (limited) time with deliberation, and sometimes they want to hear the input string a second time. But even in this (worst) case recognition happens in linear time, and in this respect at least, finite automata constitute realistic models of native speakers.

The importance of this fact for linguistic metatheory should not be underestimated: those frameworks (like transformational grammar, see Rounds 1975) that generate languages with exponential (polynomial) recognition complexity make the prediction that there are problems which can be solved both by humans and Turing machines in a measured time, with humans showing an exponential (polynomial) gain over machines in the long run. For instance, Lexical-Functional Grammar (see Bresnan 1983) makes the claim that humans can solve certain NP-hard problems in linear time (cf. Berwick 1982), and this is not very likely. On the other hand, those frameworks (like Generalised Phrase-Structure Grammar, see Gazdar 1982) that generate only languages of polynomial time complexity might have some psychological reality; at least there is nothing in present-day complexity theory that precludes the possibility of one implementation (e.g. multi-tape Turing machines, or, for that matter, the brain) gaining a small polynomial factor over another one (e.g. single-tape Turing machines).

Another advantage of Type 3 NL models is that they make the problem of language acquisition solvable, at least theoretically. It is well known that no algorithm can decide whether a context-free grammar generates a given context-free language: therefore, if every (infinite) context-free language is a possible NL, Church's thesis will make it impossible for the child to acquire one, not even in case they have access to an oracle (say, the parents) that tells them whether a string belongs to the language or not. Therefore, it is sometimes supposed that the primary linguistic data accessible during language acquisition contains not only strings, but the associated tree-structures as well. But if NLs are regular, the problem is solvable without recourse to this rather strange assumption: given an upper bound on the number of states in the canonical automaton generating the language, it is possible to reconstruct the automaton in a finite number of queries (Moore 1956). Since the number of queries is at least 10^{12} even if the child has access to a 'representative sample' of 10^4 sentences that reaches every live state in the automaton (see Angluin 1981), and it is impossible to make more than 10^6 queries in a lifetime, NLs must form a proper subset of regular languages.

In fact, there is reason to suppose that every NL will be *non-counting.* Specifically, a string xy^4z will belong to some NL if and only if xy^5z is also in it. This is obvious if y is a coordinate conjunct – as certain languages differentiate between singular, dual, trial, and plural, the

[8] This property is a linear version of the Depth Hypothesis (Yngve 1961).

[9] For the sake of simplicity I have supposed that sentences in embedded position are freely interchangeable, i.e. that there is only one accepting state.

number 4 cannot be reduced. If y is the repeated part of some left- or right- recursive construction (e.g. a *that* -clause), five copies will be just as grammatical as four copies were: the converse also seems to hold. If this characterization of NLs is true, the traditional mode of language description in 'tactical' terms is fully justified, because every non-counting language can be built up from the elements of the alphabet using only catenation and Boolean operations (McNaughton-Papert 1971). Conversely, as the traditional phonotactic, morphotactic, and syntactic descriptions of NLs used only catenation, union, intersection, and sometimes complementation (in the form of 'negative conditions'), the generative power of the Item and Arrangement model (see Hockett 1954) does not exceed that of counter-free automata.

It is also possible to develop a *lower bound* for the family of NLs: the phenomenon of syntactic concord over unbounded domains (which I suppose to be present in every NL) will guarantee that NLs cannot be *locally testable.* The following demonstration is based on a regular expression used by Pullum- Gazdar (1982): coordination (the outermost Kleene *) has been added in order to create non-initial and non-final elements that have to agree with each other.

(12) ((Which problem did your professor say ((she + you) thought)* was unsolvable) + (Which problems did your professor say ((she + you) thought)* were unsolvable))*

Suppose, indirectly, that English is k-testable for some fixed k, and consider the following strings:

(13) ((Which problem did your professor say (she thought you thought)k was unsolvable) (Which problems did your professor say (she thought you thought)k were unsolvable))2

(14) ((Which problem did your professor say (she thought you thought)k were unsolvable) (Which problem did your professor say (she thought you thought)k was unsolvable) (Which problems did your professor say (she thought you thought)k was unsolvable) (Which problems did your professor say (she thought you thought)k were unsolvable))

Apart from the order of the conjuncts, the only difference between (13) and (14) is that in the latter subject-predicate number agreement is violated in the first and the third conjuncts. Therefore, (14) is ungrammatical, but it has the same subwords of length k (and with the same multiplicity) as the grammatical (13). This contradicts our hypothesis that English was k-testable.

Hopefully, this special position of NLs in the Chomsky hierarchy can be utilized in streamlining the (oracle) algorithms modeling language acquisition, because such algorithms (if actually implemented) would greatly simplify the descriptive work of the linguist, and, at least to a certain extent, would finally fulfill the structuralists' promise of discovery procedure.

Acknowledgement

I am grateful to Gerald Gazdar for his valuable criticisms of an earlier version of this article. The usual disclaimers apply.

References

Angluin, D. 1981: A note on the number of queries needed to identify regular languages. Information & Control 51, 76-87

Arrow, K. I. 1950: A difficulty in the concept of social welfare. Journal of Political Economy 58, 328-346

Berwick, R. C. 1982: Computational complexity and Lexocal-Functional Grammar. American Journal of Computational Linguistics 8, 97-109

Bloomfield, L. 1926: A set of postulates for the science of language. Language 2, 153-164

Borgida, A. T. 1983: Some formal results about stratificational grammars and their relevance to linguistics. Mathematical Systems Theory 16, 29-56

Bresnan, J. 1983: The mental representation of grammatical relations (ed) MIT Press, Cambridge, Mass.

Carroll, J. B. 1953: The study of language. Harvard University Press, Cambridge, Mass.

Chomsky, N. 1957: Syntactic Structures. Mouton, The Hague Chomsky, N. 1956: Three models for the description of language. I.R.E. Transactions on Information Theory IT-2.

Chomsky, N. 1961: Some methodological remarks on generative grammar. Word 17, 219-239

Chomsky, N. 1963: Formal properties of grammars. In: Luce - Bush - Galanter (eds) Handbook of mathematical psychology. Wiley, New York

Chomsky, N. 1970: Remarks on nominalization. In: Jacobs-Rosenbaum (eds): Readings in English Transformational Grammar. Ginn, Waltham, Mass.184- 221

Church, K. 1980: On parsing strategies and closure. Proceedings of the 18th Annual Meeting of the

ACL 107-111
Eilenberg, S. 1974: Automata, languages, and machines. Academic Press, New York
Gazdar, G. 1982: Phrase structure grammar. In: Jacobson-Pullum (eds): The Nature of Syntactic Representation. Reidel, Dordrecht 131-186
Greenbaum, J.- R. Quirk 1970: Elicitation experiments in English: linguistic studies in usage and attitude. Longman, London
Greene, J. 1972: Psycholinguistics. Penguin, Harmondsworth
Harris, Z. 1946: From morpheme to utterance. Language, 22, 161-183
Harris, Z. 1951: Methods in Structural Linguistics. University of Chicago Press
Harrison, M. A. 1978: Introduction to Formal Language Theory.Addison- Wesley, Reading, Mass.
Hockett, C. 1954: Two models of grammatical description. Word 10, 210-231
Hockett, C. 1955: A manual of phonology. International Journal of American Linguistics, Memoire 11
Itkonen, E. 1981: The concept of linguistic intuition. In: Coulmans (ed): A Festschrift for Native Speaker. Mouton, The Hague, 127-140
Kleene, S. C. 1956: Representation of events in nerve nets and finite automata. In: Shannon - McCarthy (eds): Automata studies. Princeton University Press 3-41
Lamb, S. M. 1966: Outline of stratificational grammar. Georgetown University Press, Washington D.C.
Levi-Strauss, C. 1958: Anthropologie structurale. Plon, Paris.
Lieber, R. 1981: On the Organization of the Lexicon. IULC
Marcus, S. 1964: Grammatici si automate finite. Editure Academiei, Bucharest
Marks, L. E. 1968 Scaling of grammaticalness of self-embedded English sentences. Verbal Learning & Verbal Behavior 7, 965-967
Matthews, R.J. 1979: Are the grammatical sentences of a language a recursive set? Synthese 40, 209-224
McCullogh, W. S. - W. Pitts 1943: A logical calculus of the ideas immanent in nervous activity. Bulletin of mathematical biophysics 5, 115-133
McNaughton, R. - S. Papert 1968: The syntactic monoid of a regular event. In Arbib (ed): Algebraic theory of machines, languages, and semigroups. Academic Press, New York 297-312
McNaughton, R - S. Papert 1971: Counter-free automata. Research Monograph no. 65, MIT Press, Cambridge, Mass.
Miller, G. A. - S. Isard 1964: Free recall of self-embedded English sentences. Information & Control 7, 293-303
Moore, E. F. 1956: Gedanken-experiments on sequential machines. In: Shannon - McCarthy (eds): Automata studies. Princeton University Press 129-153
Newmyer, F. J. 1980: Linguistic theory in America. Academic Press, New York
Pullum, G. and G. Gazdar 1982: Natural languages and context free languages. Linguistics and Philosophy 4, 471-504
Putnam, H. 1961: Some issues in the theory of grammar. Proc. Symposia in Applied Mathematics 1961
Reich, P.A. 1969: The finiteness of natural languages. Language 45, 831-843
Robins, R.H. 1959: In defense of WP. Trans. Philol. Soc. 116-144
Rogers, H. 1967: The theory of recursive functions and effective computability. McGraw-Hill, New York
Rounds, W. 1975: A grammatical characterisation of exponential-time languages. Proc. 16th Symposium on Switching Theory and Automata 135-143
Salomaa, A. 1973: Formal Languages. Academic Press, New York
Sullivan, W. J. 1980: Syntax and linguistic semantics in stratificational theory. In: Wirth- Moravcsik (eds) Current approaches to syntax. Academic Press, New York 301-327
Yngve, V. H. 1961: The depth hypothesis. In: Jakobson (ed): Proc. of Symposia in Applied Mathematics 12, 130-138