SETS AND SENTENCES

D. T. LANGENDOEN AND P. M. POSTAL

1. BACKGROUND AND GOALS

LINGUISTIC research over the past quarter century has been largely guided by two major assumptions introduced by Noam Chomsky: (i) that the best theory of a natural language (NL) is a grammar that generates its sentences and (ii) that human beings know an NL in virtue of knowing that grammar. These assumptions cannot be maintained. The collection of sentences comprising each individual NL is so vast that its magnitude is given by no number, finite or transfinite. This means that NLs cannot, as is currently almost universally assumed, be considered recursively enumerable, hence countable (or denumerable) collections of sentences. For if they were such, the magnitude of each would be no greater than the smallest transfinite cardinal number $\aleph_0$. It then follows that there can be no procedure, algorithm, Turing machine or grammar that constructs or generates all the members of an NL, since, by definition, such a grammar can construct or generate only recursively enumerable, hence countable, collections. A system which constructs *some* NL sentences must inevitably leave most NL sentences unconstructed.

2. THE ANALOGY WITH CANTOR'S RESULTS

2.1 *Co-ordination*

Our conclusion concerning the vastness of NLs is based on a demonstration of a strict parallelism between the collection of all sentences of an NL and the collection of all sets. The discovery around the turn of this century that the latter collection is not itself a set led to

Adapted by the authors from *The Vastness of Natural Languages*. Reprinted by permission of Basil Blackwell Publisher.

fundamental reforms in logic and the foundations of mathematics. The same reasoning that establishes that the collection of all sets cannot itself be a set (a collection with fixed magnitude, finite or transfinite) also establishes that the collection of all NL sentences cannot be a set. Consequently, fundamental revisions of currently standard views of NLs and grammars are required for reasons similar to those that operated in the foundations of logic and mathematics.

We assume that co-ordinating particles like English *and, or*, etc., have a structure in which, quoting Gazdar (1981, p. 158):

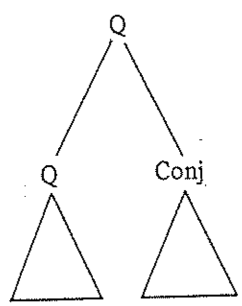
> . . . the co-ordinating word forms a constituent with the immediately following node and is not simply a sister of all the conjuncts.

However, we generalize to eliminate reference to 'following' to cover NLs where the co-ordinating particle follows. For concreteness and ease of reference, we assume these particles belong to a grammatical category called Conj, whose elements have the properties in (1):

(1) If A is a Conj node, then there exist nodes B and C such that:
  (*a*) There is a grammatical category Q such that both B and C are Q nodes; and
  (*b*) A is a daughter of B and the unique sister of C.

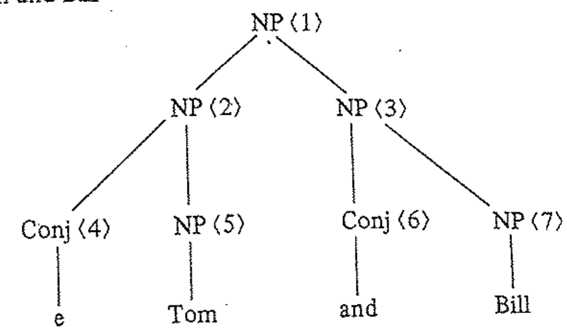
In these terms, we say that nodes instantiating variable B in (1) are *conjuncts*, while those instantiating C are *subconjuncts*. Ignoring the order of sisters, every conjunct thus has the structure in (2):

(2)



We allow a Conj terminal to be null for NLs without visible co-ordinating particles and for cases in NLs like English where one or more instances of Conj are not visible. Thus we take (3*a*) to have a structure like (3*b*), associating nodes with numbers for ease of reference:

(3) (*a*) Tom and Bill
(*b*)



In this case, according to our definitions, nodes 2 and 3 are conjuncts, while nodes 5 and 7 are subconjuncts.

Nodes like 1 in (3*b*) will be referred to as *co-ordinate compound constituents (nodes)*, definable as in C (4):

(4) A constituent (node) Q is a co-ordinate compound of grammatical category C if and only if:
  (*a*) Q is of category C; and
  (*b*) Q has at least two immediate constituents; and
  (*c*) each of Q's immediate constituents is a conjunct.

Observe that nothing in the definition of 'co-ordinate compound' imposes any upper limit, finite or transfinite, on the number of immediate constituents in such structures. Co-ordinate compounds are evidently subject to the two fundamental lawful restrictions in (5):

(5) All sister conjuncts are of the same grammatical category as
  (*a*) each other; and
  (*b*) the co-ordinate compound constituent of which they are immediate constituents.

However, we take C in (4) to vary only over so-called 'major' categories, so as not to exclude the possibility of, e.g. compounds of the category Plural with exclusively singular immediate constituents, etc.

Co-ordinate compounding in NLs is, we claim, governed by a fundamental condition we refer to as *co-ordinate compound closure*. To characterize this principle it is convenient to introduce two more basic terms, *co-ordinate projection*, and *projection set* as in (6):

(6) Let U be a set of constituents all of category Q and of cardinality > 1 and let T be some co-ordinate compound of category Q. Furthermore, assume that:

(*a*) each conjunct of T has an element of U as a subconjunct; and

(*b*) each element of U is a subconjunct of some conjunct of T; and

(*c*) no element of U appears more than once as a subconjunct of any conjunct of T; and

(*d*) if two elements of U occur as subconjuncts of conjuncts C_i and C_j of T, then C_i and C_j occur in a fixed order. Where C_i and C_j are of distinct lengths, assume the shorter precedes; where C_i and C_j are the same length, assume some arbitrary order.

In this case, we say both that T is a *co-ordinate projection of U* and that U is *the projection set of T*.

(*e*) Example: Let U = {Laura, Maxine, Brooks}. Then one may choose T = *Laura, Maxine, and Brooks*.

(6*a*) demands that the subconjuncts for T be drawn exclusively from U, while (6*b*) demands that each element of U be used to form some subconjunct. (6*c*) prevents repetitions of elements from the projection set. (6*b,c*) together determine *inter alia* that the number of conjuncts in a co-ordinate projection is identical to the number of elements in its projection set. (6*d*) insures that different orders of conjuncts are irrelevant. Given the latter, co-ordinate projections of a set of constituents are unique *up to choice of element of Conj*. For simplicity, the discussion is henceforth limited to some unique element of the category Conj, thus determining a unique co-ordinate projection for any set of constituents, making it sensible to speak of *the* co-ordinate projection of such a set.

As stressed by E. Keenan (personal communication), it is important to show that every subset of a collection of constituents of category Q has a co-ordinate projection. But this is straightforward. For consider some such subset U. Take the cardinality of U to be *k*, with *k* indifferently finite or transfinite. Clearly, from the purely formal point of view, there is a co-ordinate compound W belonging to the category Q, hence having immediate constituents of the category Q. Moreover, each of these immediate constituents is a conjunct. Since there are no size restrictions on co-ordinate compounds, W can have any number, finite (> 1) or transfinite of immediate constituents. W can then, in particular have exactly *k* such constituents. Each of these (conjuncts) has one and only one subconjunct. The set of all such subconjuncts, call it V, obviously then also has exactly *k* members. To show that W is a co-ordinate projection of U, it then in effect suffices that there exist a

one-to-one mapping from U to V. But this is trivial, since the two sets have the same number of elements. However, the conclusion that each subset U of constituents of category Q has a co-ordinate projection does not mean that the co-ordinate projection is necessarily well-formed in the NL from which U is drawn. The latter can only be determined by axioms to this effect, to which we now turn.

The notion of closure for co-ordinate compounding is stated in (7):

(7) The Closure Principle for Co-ordinate Compounding
If U is a set of constituents each belonging to the collection, S_W, of (well-formed) constituents of category Q of any NL, then S_W contains the co-ordinate projection of U.

Crucially, principle (7) has a 'recursive' property in that it also refers to cases where members of U are themselves co-ordinate compounds.

Although it is not entirely clear for what categories principle (7) holds universally, it is, we claim, at least valid for the one case of real concern here: where Q is the category S(entence). This yields (8), which we take as a truth about all NLs:

(8) Closure under Co-ordinate Compounding of Sentences

(*a*) If U is a set of constituents each belonging to the collection, S_W, of (well-formed) constituents of the category S of any NL, then the co-ordinate projection of U belongs to S_W.

More precisely, (8*a*) can be stated as in (8*b*):

(*b*) Let L be the collection of all members of the category S of an NL and let CP(U) be the co-ordinate projection of the set of sentences U.
Then:

$$(\forall U)(U \subset L \rightarrow CP(U) \in L)$$

Principle (8) is no doubt too general in one respect. Sentences (clauses) fall into types, declarative, interrogative, imperative, etc. And a co-ordinate compound is in general only freely permitted for members of a single type. One could amend (8) appropriately, by restricting the members of U to a single type. We ignore this complication in the wording of what follows.

The principle of closure for co-ordinate compounding of sentences formalizes the following observation about collections of attested NL sentences. Given any set of Ss (of the same type) of some NL L, there is a well-formed co-ordinate compound of those Ss in L, as illustrated in (9), where a double arrow means that the sentence on its right is the co-ordinate projection of the set of sentences on its left.

(9) (*a*) {Gregory is handsome; It is raining; Figs can kill} $\Longrightarrow$
(*b*) Gregory is handsome, it is raining and figs can kill.
(*c*) {Gregory is handsome; It is raining; Figs can kill; Gregory is handsome, it is raining, and figs can kill} $\Longrightarrow$
(*d*) Gregory is handsome, it is raining, figs can kill and Gregory is handsome, it is raining and figs can kill.

We now show that the closure principle for co-ordinate compounding leads to the conclusion that there is no set of NL sentences just as the axiom of powers leads to the conclusion that there is no set of all sets.

2.2. *The Cantorian Analogue*

Let L be an NL whose ordinary vocabulary contains the name *Z* of a particular person or elephant. Assume L contains a denumerably infinite set, S_0, of *noncompound* sentences, each of which is about the entity Z, named by *Z*. This assumption seems uncontroversial, since, for many known NLs, it is easy to effectively specify such a set. For example, if L is English, S_0 could be the set in (10), where Z = Babar.

(10) {Babar is happy; I know that Babar is happy; I know that I know that Babar is happy; I know that I know that I know that Babar is happy . . . }

Assume that L is closed under co-ordinate compounding of clauses, that is, obeys (8). Then L also contains a set S_1 made up of all the sentences of S_0 together with all and only the co-ordinate projections of every subset of S_0 with at least two elements, that is, with a set containing one co-ordinate projection for each member of the power set of S_0 whose cardinality is at least 2. The clumsiness of this formulation arises from the fact that co-ordinate projections, given the nature of co-ordination, require by definition at least two subconjuncts, while power sets contain singletons as well as the null set. To simplify the discussion, we utilize the notation *>1-power set of X*, meaning that proper subset of the power set of the set X containing all and only the power set elements of cardinality 2 or greater. To illustrate, if S_0 is as in (10), then S_1 can be taken as the set in (11):

(11) {Babar is happy; I know that Babar is happy; I know that I know that Babar is happy; . . . ; Babar is happy and I know that Babar is happy; Babar is happy and I know that I know that Babar is happy; . . . ; Babar is happy, I know that Babar is happy, and I know that I know that Babar is happy; . . . }

By assumption, the cardinality of S_0 is $\aleph_0$. To determine the cardinality of S_1, one can appeal directly to Cantor's theorem. Each member of S_1 can be put into one-to-one correspondence with a non-null member of the power set of S_0, determined as follows. Each non-compound sentence of S_1 corresponds to the singleton set whose unique element is the corresponding sentence of S_0. Each co-ordinate compound sentence of S_1 corresponds to its projection set. Hence each compound sentence of S_1 with two conjuncts corresponds to the set made up of the corresponding pair of sentences of S_0, each compound sentence of S_1 with three conjuncts corresponds to the set made up of the corresponding triple of sentences of S_0. Similarly, for each finite subset of S_0 of cardinality >3, there is a corresponding compound sentence of S_1, namely, the co-ordinate projection of that subset of S_0. Finally, each *infinite* subset of S_0 also corresponds to a compound sentence of S_1, although, of course, each such co-ordinate projection *is of transfinite length*. Overall then, each co-ordinate compound sentence of S_1 corresponds to a member of the >1-power set of S_0. Since the cardinality of the power set of any denumerably infinite set, and hence of S_0, is of the order of the continuum, that is $\aleph_1$, the cardinality of S_1 is $\aleph_1$. Further, since L is closed under co-ordinate compounding, the sentences of S_1 are all contained in L, and therefore, if L has any determinate magnitude, this must be of *at least* the cardinality $\aleph_1$.

The set S_1 as a whole is characterizable as in (12), where '$\cup$' is the sign for set union.

(12) $S_1 = S_0 \cup K_0$,

where $K_0 = \{x: (\exists y)(y \subseteq S_0 \wedge x \text{ is the co-ordinate projection of } y)\}$

In other words, S_1 is the union of S_0 and the set K_0 consisting of all and only the co-ordinate projections of the >1-power set of S_0.

The cardinality of S_1 exceeds that of S_0 precisely because it contains sentences with transfinitely many co-ordinated constituents. The cardinality of the set S_0, made up of the union of S_0 with all those sentences of S_1 with at most finitely many conjuncts as immediate constituents is also $\aleph_0$. But the set S_1', the union of all of the sentences of S_0' together with the co-ordinate compound sentences of L whose immediate constituents are conjuncts with only sentences of S_0 as sub-conjuncts, is of the cardinality $\aleph_1$.

If English is governed by (8) and contains the sentences in S_0, then it also contains the sentences in S_1. That is, English contains at least as many sentences as the continuum.

But it evidently must contain even more. Consider the union of S_1 and a set containing the co-ordinate projection of every member of the >1-power set of S_1. That is, consider the set S_2, definable analogously to S_1 in (12), as in (13):

(13) $S_2 = S_1 \cup K_1$,

where $K_1 = \{x: (\exists y)(y \subseteq S_1 \wedge x \text{ is the co-ordinate projection of } y)\}$

Via the procedure outlined for S_1, the members of S_2 can be put into a one to one correspondence with the members of the power set of S_1, excluding the null set. Hence the cardinality of $S_2 = \aleph_2$. Futher, since L is closed under co-ordinate compounding, S_2 is also included within L. Consequently, the magnitude of L, if determinate, is at least of the cardinality $\aleph_2$.

Just as Cantor showed for power sets in general, the possibility of forming greater and greater sets of NL sentences always remains. For any set of sentences like S_1, S_2, etc., there is always a still bigger set *included in L*, given by the schematic characterization in (14):

(14) $S_i = S_{i-1} \cup K_{i-1}$, where $i > 0$ and where

$K_{i-1} = \{x: (\exists y)(y \subseteq S_{i-1} \wedge x \text{ is the co-ordinate projection of } y)\}$

At no point can a set of sentences be obtained that exhausts an NL having sentence co-ordination governed by the closure law (8). Naturally, this will not be less true if one begins, more realistically, with *all* of the finite sentences of that NL, not just an artificially small subset of these like (11) containing only expressions sharing a single name. To prove that no set of sentences can exhaust an NL, it suffices to construct an analogue of Cantor's Paradox *from the contrary assumption*, a construction which the previous remarks make directly possible. We call this result *the NL Vastness Theorem*, and state it as (15):

(15) *The NL Vastness Theorem*

THEOREM: NLs are *not* sets (are megacollections)

Proof:

Let #X be the cardinality of an arbitrary set X and let L be the collection of all sentences of some NL.

(*a*) Assume to the contrary of the theorem that L is a set.

(*b*) Then L has a fixed cardinality, #L.

(*c*) Since L is closed under co-ordinate compounding, L contains a subset consisting of all and only the co-ordinate projections of the >1-power set of L. Moreover, each member of the >1-power set of L has a co-ordinate projection. Hence $(\exists Z)(Z \subseteq L)$, where:

$Z = \{x: (\exists y)(y \subseteq L \wedge x \text{ is the co-ordinate projection of } y)\}$

(*d*) Since many sentences in L, in particular, all those elements of L which are not co-ordinate compounds, are *not* in Z, Z is a *proper* subset of L. That is, not only $Z \subseteq L$ but in fact $Z \subset L$.

(*e*) Hence, $\#Z \leq \#L$.

(*f*) But #Z is, given the definition of Z in (15*c*), of the order of the power set of L.

(*g*) Hence, by Cantor's Theorem, $\#Z > \#L$.

(*h*) Since conclusion (15*g*) contradicts conclusion (15*e*), assumption (15*a*) is false.

The assumption that L is a set, hence a collection having a fixed cardinality, yields a contradiction and is thus necessarily false. Therefore, the collection L is not a set. But L in (15) was arbitrarily chosen. Just as Cantor's Paradox shows there is no single set containing all non-null sets, the NL Vastness Theorem shows that an NL can be identified with no fixed set of sentences at all, no matter how great its cardinality. Like the collection of all sets, an individual NL must be regarded as a megacollection.

2.3 *The Mathematical Argument as a Linguistic Argument*

Having constructed the central argument of this study, we now comment on its character. The demonstration in (15) that NLs are not sets but megacollections has, like any attempt to apply a mathematical result to some domain of facts, two distinguishable aspects. There must, first, be a proof of the relevant theorem, a question of formal mathematics, involving a purely demonstrative argument and, second, an argument, in general necessarily nondemonstrative, that the relevant domain of facts manifests all crucial properties of the mathematical assumptions underlying the proof of the theorem. In this case, the relevant theorem is the NL Vastness Theorem, whose proof corresponds closely to the proof of Cantor's Paradox. The second aspect, the consequence that this formal proof 'applies' to NLs, involves the claim that NLs do indeed model a system of mathematical objects having the properties which yield the NL Vastness Theorem. Only by confirming the second aspect of the argument can one avoid the problem properly noted by Hockett (1966, p. 186):

> An ironclad conclusion about a certain set of 'languages' (in the formal sense) can be mistaken for a discovery about real human language.

There is another way to put the point. As with any proof from assumptions A to a conclusion Z, one can regard the NL Vastness Theorem as a proof of the conditional $A \Rightarrow Z$. This proof does not require that A be true. But the detachment of Z as a true consequence then only follows via Modus Ponens, which requires that the antecedent of a conditional be true. Therefore, (15) is a proof of a conditional whose consequent is the conclusion that NLs are megacollections. But to derive the actual nonconditional conclusion, that is, the NL Vastness Theorem itself, via Modus Ponens requires that the antecedent to be true. In effect, this antecedent is the claim that NL co-ordination is governed by the closure principle (8). Surely, scepticism about the NL Vastness Theorem must focus on this axiom, which is not a traditionally or currently accepted linguistic principle.

Let us therefore briefly refocus attention on condition (8), the claim that NLs are closed under the co-ordinate compounding of sentences. Although not a familiar principle of past or present linguistics, (8) expresses, we claim, a profound truth about NLs. It says not only that the principles of grammatical theory and the rules of grammar directly relevant to characterizing co-ordinate structures must not themselves preclude closure, but that no other rules can have this effect either. No matter how one characterizes the collection of co-ordinate structures of English, closure would be violated if some *independent* English rule said, for example, that there was a maximum bound on number of conjuncts, or one which said that some particular pair of clauses of the same type could not form a co-ordinate compound, etc. Similarly, (8) would be violated if some rule of English required every co-ordinate compound to have *more* than k conjuncts for some fixed k, or if there were a rule ('filter') precluding, e.g. the sequence of English words *and+it*. But the known facts about co-ordinate compounding in NLs reveal the existence of no such constraints. Principle (8) claims that the lack of such is nonaccidental.

Closure principle (8) plays a role in the proof of the NL Vastness Theorem analogous to that played in set-theoretical discussions (in particular, the proof of Cantor's Paradox) by axioms which determine that every set does have a power set. Such axioms guarantee that the collection of sets is closed under power setting in essentially the way principle (8) guarantees that the collection of sentences of an NL is closed under co-ordinate compounding. It seems that there are exactly as good grounds for the latter as for the former.

Principle (8) mentions a set U of constituents but says nothing about its magnitude. Clearly, one obtains a variety of different closure laws by imposing differential magnitude requirements on U, as in (16):

(16) (*a*) U has less than k elements (k a positive integer).
(*b*) U has less than $\aleph_0$ elements.
(*c*) U has less than $\aleph_1$ elements.
(*d*) U has less than $\aleph_2$ elements.

There are infinitely many possible magnitude restrictions on U, each limiting the collection of possible projection sets for co-ordinate compounds. If any of these are adopted *instead of (8)*, the argument that NLs are not sets will obviously not go through, because at some point in the definitions of sets S_3, S_4, etc., schematized in (14), the resulting co-ordinate compounds will not be determined to be included in the language.

More precisely, if one of the denumerably many restrictions in (16*a*) is chosen, the collection of co-ordinate compounds is not determined to be more than a finite set, while if (16*b*) is chosen, it is a countably infinite set. Consequently, it is critical for the conclusion that (8) rather than any element of (16) is the *correct* closure principle for co-ordinate compounding. In particular, it is critical to justify (8) against (16*b*).

First, (8) is simpler than any statement in (16), because, unlike those statements, (8) *says nothing at all about magnitude*. Hence (8) is, by Occam's Razor, theoretically preferable to any formulation covered by (16), since it is always simpler not to specify anything about the magnitude of some collection than to say something about its size. And this obviously holds for U in (8).

Second, one can regard grammars and grammatical theory as concerned with *projecting* from the properties of *attested* NL sentences, the basic data of grammatical investigation, to the maximal lawfully characterized collections of which these attested sentences are accidental examples. One wants, given a sample of English sentences, to characterize the collection of all English sentences, and, given a sample of NL sentences, the collection of NL sentences *per se*. General scientific principles demand that the projections from the small finite samples to the desired characterizations involve the maximally general (i.e. strong) laws (principles) projecting the regularities found in observed cases to the collections as wholes. One can never justifiably replace a stronger or more general projection by a weaker or less general one unless this is factually motivated, in particular, by the excess generality leading to some false entailment, e.g. a false claim about attested examples, some contradiction, etc.

For example, there is no basis for not projecting from attested sentences of various lengths to the maximally general view that sentences of any length whatever are possible, unless this yields some false entailment, which has never been shown. Therefore, there is no basis for not projecting from attested co-ordinate compounds of various lengths to the maximally general view, represented by (8), that co-ordinate compounds of any length whatever are possible, unless this yields some false entailment, which again has not been shown.

Thus, there are two reasons for choosing the closure principle (8) over any of the alternatives in (16): (8) is both simpler and stronger.

Obviously, the conclusion which (8) determines, that NLs are megacollections, is itself no basis whatever for rejecting this principle, any more than the conclusion which the Axiom of Powers determines, that the collection of all sets is a megacollection, is a ground for rejecting that axiom. Essentially, principle (8) says that it makes no more sense to think that structures otherwise having the structural (linguistic) properties of co-ordinate compounds nonetheless fail to be co-ordinate compounds if they have more than some fixed number of conjuncts than it does to think that aggregates fail to be sets if they have more than some fixed number of elements. That is, it is as arbitrary to claim that some structures have too many conjuncts to be proper co-ordinate compound sentences as it is to claim that some aggregates have too many elements to be (power) sets.

To sum up, (8), the principle of closure under co-ordinate compounding, plays an absolutely crucial role in the argument that NLs are megacollections. More precisely, it is the critical assumption guaranteeing that NLs are models of a system of objects for which all the mathematical assumptions underlying the proof of the NL Vastness Theorem hold.

The argument given in (15) involves the existence of sentences of transfinite length, the postulation of which, of course, clashes with standardly held but unmotivated and never justified views. The standard view is that while there is no longest sentence, every sentence is of finite length, that is, has a length less than $\aleph_0$. This amounts to imposition of what we will call *a length law* on NL sentences. Our claim is that no such length law is true of NL sentences. It is crucial, moreover, that the nonexistence of a length law is *not* a premise of the proof in (15) but, rather, is a corollary of the closure principle (8). This is shown in (17), which demonstrates the nonexistence of any length law for NL sentences, finite or transfinite. The proof utilizes a predicate Length, taken to be a measure of the number of words in a sentence. We also make use of the self-evident fact that the length of any co-ordinate projection is not less than the cardinality of its projection set. This is only to say that each member of a projection set T contributes at least one word to the co-ordinate projection of T.

(17) *The No Upper Bound Theorem*
THEOREM: Let L be the collection of all sentences of some NL. Then:

$$(\forall k)(\text{Cardinal}\,(k) \Rightarrow (\exists x)(x \in L \wedge \text{Length}(X) \geqslant k)).$$

Proof:

(*a*) Assume to the contrary that j is a cardinal such that:

$$(\forall Y)(Y \in L \Rightarrow \text{Length}(Y) < j).$$

(*b*) Every proper subset of L then has a cardinality $< j$. For the closure axiom (8) determines that every such subset is the projection set of some co-ordinate projection which is a sentence of L. And, as we have seen, the length of any co-ordinate projection is at least that of the cardinality of its projection set. Hence if some $C \subset L$ had $> j$ members, some $Z \in L$ would have a length $> j$, namely, for a Z equal to a co-ordinate projection of C.

(*c*) We now show that if every proper subset of L has a cardinality $< j$, the maximal cardinality of L is j. There are two cases, since L is either finite or not finite.

(i) Case A. L is finite. Consider one member, M, of the set of *biggest* proper subsets of L. M will have one less member than L. Since M has, from (*b*), a cardinality $< j$, the maximum cardinality of L is j.

(ii) Case B. L is transfinite. It follows from set theory that L is equipollent to some proper subset of L, call it D. Since, from (*b*), D has a cardinality $< j$, so does L.

(*d*) It follows from (*c*) that L is a set with $\leqslant j$ members, contradicting the NL Vastness Theorem. Hence (*a*) is false.

The consequence that NLs are megacollections rather than recursively enumerable sets cannot be rationally avoided by a decision to adopt the finiteness limitation on sentence size or its analogue for the number of conjuncts even in the absence of substantively or logically motivated bases for such conditions. We are rejecting an argument which might go something like (18).

(18) The finiteness limitations are justified just because they subsume NLs within the realm of recursively enumerable sets and Turing machine grammars, a mathematically well-understood domain about which a rich, useful body of knowledge has been accumulated.

The fallacy in such a defence of a closure principle like (16*b*) has already in effect been uncovered by Chomsky several times in different contexts. First, consider (19):

(19) Chomsky (1957: 23):

We might arbitrarily decree that such processes of sentence formation in English as those we are discussing cannot be carried out more than *n* times, for some fixed *n*. This would of course make English a finite state language, as, for example, would a limitation of English sentences to length of less than a million words. *Such arbitrary limitations serve no useful purpose, however*.' [Emphasis ours: DTL/PMP.]

While Chomsky's comment about 'fixed *n*' was intended only to cover finite instantiations of *n*, the force of the remarks clearly carries over to his own choice of length law and all others as well, since these are nothing but instances where *n* varies over transfinite cardinals. The same point applies to (16*b*).

Again, criticizing a certain argument which need not concern us, Chomsky made the correct observation in (20):

(20) Chomsky (1977*a*: 174):

In the first place, he is overlooking the fact that we have certain antecedently clear cases of language as distinct from maze running, basket weaving, topological orientation, recognition of faces or melodies, use of maps, and so on. We cannot arbitrarily decide that 'language' is whatever meets some canons we propose. Thus we cannot simply stipulate that rules are structure-independent, . . .

Since NLs are independently given, they are not subject to arbitrary decisions about sentence length or any other property. Just as one cannot simply decide that rules are (or are not) structure-dependent, one cannot just decide that sentences are (or are not) all finite, or that the number of conjuncts in a co-ordinate compound is always finite. In both cases, arguments based on the nature of the attested part of the subject matter are required. Consequently, one can no more decide that each sentence is finite in length than one can decide that each is less than one thousand morphemes in length or that each

grammar is a finite state system. Unfortunately for linguistics, the sentence finiteness decision has been *arbitrarily* made and maintained for nearly thirty years. But this past mistake contains no justification for its continuation.

3. IMPLICATIONS

3.1 *Remarks*

So far we have established the two relatively simple substantive points about NLs in (1):

(1) (*a*) The NL Vastness Theorem; that is, the existence of unbounded co-ordination subject to the closure principle (8) of Section 2 entails, via a Cantorian analogy, that the collection of sentences in NL is (i) bigger than countably infinite, and (ii), in fact, a megacollection.

(*b*) The No Upper Bound Theorem, that is; there is no length law on NL sentences.

Moreover, we showed that (1*b*) is a logical consequence of the closure principle, which thus provides a principled reason for the nonexistence of NL length laws. Since the argument for (1*a*) was based exclusively on English data, it is more accurate to say that (1*a*) follows for any NL manifesting co-ordination with the essential properties characterized earlier.

However, we know of no NL ever described which has even been *claimed* to lack co-ordination of, *inter alia*, clauses, as expressed by Dik:

(2) Dik (1968: 1):

For a variety of reasons the so-called 'coordinative construction' is of special importance to linguistic theory. In the first place, this type of construction seems to be a universal feature of natural languages. Secondly, not only does its existence seem to be universal, but the way in which it is manifested in each particular language also shows a quite general, if not universal pattern.

Consequently, we hypothesize that both (1*a*, *b*) are proper universal truths about NLs.

3.2 *Linguistic Consequences*

(1) can be used, as earlier discerned facts about the nature of NLs have been, to falsify proposed grammatical theories on the grounds that

they are too weak. Just as certain facts about NLs were taken to show that finite state grammars, context-free grammars, etc., are too weak, the fact that NLs are megacollections shows that any conception of NL grammars under which they are Turing machines is inadequate. Put differently, (1*a*) entails that any theory which claims NLs are recursively enumerable sets is false. We formulate this consequence explicitly as a theorem referred to as *the NL Nonconstructivity Theorem*, given in (3):

(3) THEOREM: No NL has any constructive (= proof-theoretic, generative or Turing machine) grammar

Proof:

(*a*) Let L be an NL and let G be a constructive grammar.
(*b*) G specifies exactly some collection, call it C(G). From the definition of constructive systems, G recursively enumerates C(G), which is hence a countably infinite or finite set.
(*c*) The NL Vastness Theorem shows that L is a megacollection.
(*d*) Thus (Eisenberg (1971: 304)) L > C(G); and hence C(G) ≠ L.
(*e*) Therefore, G is not a grammar of L.

Since G and L in (3) were arbitrarily chosen, it has been shown that *no* constructive system is a correct grammar of *any* NL.

Although the NL Nonconstructivity Theorem is straightfoward, its consequences are both extraordinarily broad and deep. For, as Chomsky observed in the passage in (4):

(4) Chomsky (1957: 34):

The strongest possible proof of the inadequacy of a linguistic theory is to show that it literally cannot apply to some natural language.

In Chomsky's terms then, the NL Nonconstructivity Theorem shows that every variant of every view taking NL grammars to be constructive devices is a *false* theory of NLs. This means that every logically possible variant (not only those *so far* described) of all the framworks in (5) are false:

(5) *Frameworks Falsified as Theories of NLs by the NL Nonconstructivity Theorem*

Finite Grammar (Hockett (1968))
Finite State Grammar (Reich (1969))
Phrase Structure Grammar (Harmon (1963); Gazdar (1981, 1982))

Lexical/Functional Grammar (Bresnan (1982))
Realistic Grammar (Brame (1979))
Stratificational Grammar (Lamb (1966); Lockwood (1972))
Tagmemics (Longacre (1964))
Montague Grammar (Partee (1975, 1976); Dowty (1978, 1982))
Natural Generative Grammar (Bartsch and Vennemann (1972))
Semantically Based Grammar (Chafe (1970*a*, 1970*b*))
Functional Grammar (Dik (1978, 1980))
Daughter Dependency Grammar (Hudson (1976); Schachter (1978))
Phrasal Core Grammar (Keenan (1980*a*, 1980*b*))
Transformational Grammar (Chomsky (1957, 1965, 1977*b*, 1981, 1982))
Corepresentational Grammar (Kac (1980))
Relationally Based Grammar (Johnson (1979))
Dependency Grammar (Hays (1964))
Categorial Grammar (Lambek (1961))
Cognitive Grammar (Lakoff and Thompson (1975))
Meaning-Text Models (Melčuk (1981))
The Abstract System (Harris (1968))
Configurational Grammar (Koster (1981))
Neostructural Grammar (Langendoen (1982))
Augmented Transition Networks (Woods (1970))
String Adjunct Grammar (Joshi, Kosaraju and Yamada (1972))
Equational Grammar (Sanders (1972))
Systemic Grammar (Hudson (1971))
Any *constructive* system distinct from all of the preceding frameworks.

The NL Nonconstructivity Theorem actually follows from a conclusion *infinitely weaker* than the NL Vastness Theorem, namely, just from the fact that NLs are at least of the magnitude of the continuum, which suffices to justify line (d) in the proof of the NL Nonconstructivity Theorem with no reference to megacollections. Hence the stage of the analogy in (12) of Section 2 involving just the set there called S_1 already suffices to falsify all views limiting grammars to the characterization of recursively enumerable sets. This means the conclusion follows from the existence of sentences of no greater than denumerably infinite length.

The NL Nonconstructivity Theorem states that no NL has a constructive grammar. It might be wondered to just what extent the NL Vastness Theorem is incompatible with constructivity. In particular, given the syntactic nature of the data on which the NL Vastness Theorem is based, one might assume that the result was compatible, for example, with either or both constructive phonology and/or (interpretive) semantics. But this is not the case. A grammar capable of characterizing transfinite sentences cannot contain constructive phonological or semantic components.

The implications of the NL Nonconstructivity Theorem can be summed up as follows. Since the ideas of generative grammar became dominant in the late 1950s, linguistics has in general assumed that the task of grammatical theory involves answering the question: what is the right form of *generative* grammar for NLs? The many disputes which have divided linguists over the past quarter century are then reducible by and large to disputes over claims about 'right form'. Some linguists have believed that NL grammars contain transformational rules; others have denied this. Some linguists have believed that transformational rules are parochially ordered; others have denied this. Some linguists have believed that there are interpretive semantic rules; others have denied this. And so on. Underlying all such disputes has been the assumption that it is possible through appeal to some combination of proof-theoretical devices to construct *some* generative grammar for each NL. But this assumption is falsified by the NL Nonconstructivity Theorem.

There is another way to characterize the consequences summed up in (1) and in the NL Nonconstructivity Theorem. The false finiteness limitation on sentence size determined the claim that NLs fall somewhere in the domain of objects characterizable by what one might call *theoretical computer science*. Their grammars would be some sort of Turing machine, their sentence aggregates recursively enumerable sets. Since NLs are subject to no length law, they do not lie within this limited class of mathematical objects. While this conclusion may, for various socio-historical reasons, be displeasing to some, it involves no unsurmountable theoretical or methodological difficulties. Logic and the foundations of mathematics faced similar problems at the beginning of this century but did not cease to thrive; quite the contrary. Hence the results in (1) are not at all to be seen as negative or unhappy consequences for grammatical study. They can be interpreted quite positively, as showing that NLs have a grandeur not previously recognized.

The NL Nonconstructivity Theorem shows that NLs do not have generative grammars. This is quite distinct from the claim, which we totally reject, that NLs do not have explicit grammars. This is important to recognize in view of the widespread confounding of the notions 'generative grammar' and 'explicit grammar', a confounding seen in such remarks as those by Harman in (6) and Chomsky in (7).

(6) Harman (1982: vii):

> The term 'generative' derives from mathematics, not psychology. It connotes explicitness of rules, not a psychological process of sentence production. A generative grammar would therefore be a precise and explicit statement of the rules of grammar of a particular natural language like English.

(7) Chomsky (1965: 4):

> A grammar of a language purports to be a description of the ideal speaker-hearer's intrinsic competence. If the grammar is, furthermore, perfectly explicit—we may (somewhat redundantly) call it a *generative grammar*.

Contrary to the implications of such remarks, explicitness and generativeness are distinct notions. A grammar *per se* merely states necessary and sufficient conditions for membership in an NL. A *generative* grammar is, as indicated by Chomsky himself many times, not only an explicit statement of such conditions, but a procedure for enumerating the members of an NL, hence a type of Turing machine. What the NL Nonconstructivity Theorem shows, then, is that NLs have no generative grammars; but this says nothing about the possibility of nongenerative (nonconstructive) grammars of NLs. Only the confounding of the notions 'explicit grammar' and 'generative grammar' could yield the illegitimate conclusion that the NL Nonconstructivity Theorem implies that NLs do not have grammars.

Moreover, not only are nongenerative grammars a logical possibility, a substantive proposal for such exists in the literature, the nonconstructive conception of grammars in Johnson and Postal (1980) and Postal (1982). To our knowledge, this is the only *extant* view of grammar and grammatical rule which survives the NL Nonconstructivity Theorem. In this view, each grammatical rule is a statement, a formula to which truth values can be assigned and a grammar is equivalently either a set of such rules or a single logical conjunction of such rules.

3.3 *Philosophical Consequences*

The chief philosophical consequence of the preceding discussion concerns the ontological status of NLs. As in other areas, one can

The NL Nonconstructivity Theorem states that no NL has a constructive grammar. It might be wondered to just what extent the NL Vastness Theorem is incompatible with constructivity. In particular, given the syntactic nature of the data on which the NL Vastness Theorem is based, one might assume that the result was compatible, for example, with either or both constructive phonology and/or (interpretive) semantics. But this is not the case. A grammar capable of characterizing transfinite sentences cannot contain constructive phonological or semantic components.

The implications of the NL Nonconstructivity Theorem can be summed up as follows. Since the ideas of generative grammar became dominant in the late 1950s, linguistics has in general assumed that the task of grammatical theory involves answering the question: what is the right form of *generative* grammar for NLs? The many disputes which have divided linguists over the past quarter century are then reducible by and large to disputes over claims about 'right form'. Some linguists have believed that NL grammars contain transformational rules; others have denied this. Some linguists have believed that transformational rules are parochially ordered; others have denied this. Some linguists have believed that there are interpretive semantic rules; others have denied this. And so on. Underlying all such disputes has been the assumption that it is possible through appeal to some combination of proof-theoretical devices to construct *some* generative grammar for each NL. But this assumption is falsified by the NL Nonconstructivity Theorem.

There is another way to characterize the consequences summed up in (1) and in the NL Nonconstructivity Theorem. The false finiteness limitation on sentence size determined the claim that NLs fall somewhere in the domain of objects characterizable by what one might call *theoretical computer science*. Their grammars would be some sort of Turing machine, their sentence aggregates recursively enumerable sets. Since NLs are subject to no length law, they do not lie within this limited class of mathematical objects. While this conclusion may, for various socio-historical reasons, be displeasing to some, it involves no unsurmountable theoretical or methodological difficulties. Logic and the foundations of mathematics faced similar problems at the beginning of this century but did not cease to thrive; quite the contrary. Hence the results in (1) are not at all to be seen as negative or unhappy consequences for grammatical study. They can be interpreted quite positively, as showing that NLs have a grandeur not previously recognized.

The NL Nonconstructivity Theorem shows that NLs do not have generative grammars. This is quite distinct from the claim, which we

totally reject, that NLs do not have explicit grammars. This is important to recognize in view of the widespread confounding of the notions 'generative grammar' and 'explicit grammar', a confounding seen in such remarks as those by Harman in (6) and Chomsky in (7).

(6) Harman (1982: vii):

> The term 'generative' derives from mathematics, not psychology. It connotes explicitness of rules, not a psychological process of sentence production. A generative grammar would therefore be a precise and explicit statement of the rules of grammar of a particular natural language like English.

(7) Chomsky (1965: 4):

> A grammar of a language purports to be a description of the ideal speaker-hearer's intrinsic competence. If the grammar is, furthermore, perfectly explicit—we may (somewhat redundantly) call it a *generative grammar*.

Contrary to the implications of such remarks, explicitness and generativeness are distinct notions. A grammar *per se* merely states necessary and sufficient conditions for membership in an NL. A *generative* grammar is, as indicated by Chomsky himself many times, not only an explicit statement of such conditions, but a procedure for enumerating the members of an NL, hence a type of Turing machine. What the NL Nonconstructivity Theorem shows, then, is that NLs have no generative grammars; but this says nothing about the possibility of nongenerative (nonconstructive) grammars of NLs. Only the confounding of the notions 'explicit grammar' and 'generative grammar' could yield the illegitimate conclusion that the NL Nonconstructivity Theorem implies that NLs do not have grammars.

Moreover, not only are nongenerative grammars a logical possibility, a substantive proposal for such exists in the literature, the nonconstructive conception of grammars in Johnson and Postal (1980) and Postal (1982). To our knowledge, this is the only *extant* view of grammar and grammatical rule which survives the NL Nonconstructivity Theorem. In this view, each grammatical rule is a statement, a formula to which truth values can be assigned and a grammar is equivalently either a set of such rules or a single logical conjunction of such rules.

3.3 *Philosophical Consequences*

The chief philosophical consequence of the preceding discussion concerns the ontological status of NLs. As in other areas, one can distinguish three basic ontological positions potentially relevant to an account of NLs: the nominalist position identifies sentences with physical manifestations and thus cannot countenance the existence of more sentences than there are, for example, subatomic particles in the universe; the conceptualist position identifies sentences with some sort of psychological reality, for example, a mentally instantiated grammar that generates them and the realist/Platonist position takes sentences to be abstract objects, whose existence is independent of both the physical and the psychological realm.

The now standard observation that NLs are not smaller than countably infinite already drives the nominalist to the extreme view that the physical universe is infinite. But the proof that NLs are megacollections leaves the nominalist devoid of any interpretation for sentencehood. The conceptualist viewpoint tries to adapt to the infinitude of sentences by postulating an internalized, mentally real, algorithm (called a generative grammar) for constructing, in principle, each of the countably infinite number of finite sentences. This position is already problematic in that it does not assign any clear ontological status to most sentences, namely, those which are too big to be mentally constructed or to have actual mental representations. The question to be faced here is whether the conceptualist position claims that the latter sentences are real. If they are not real, what is the point of having a device which characterizes them? And if they are real, how does their reality differ from that of the realist's abstract objects? As far as we can tell, it does not, since these putatively mental objects have no physical, temporal, or psychological locus. The conclusion that NLs are megacollections simply worsens the already problematic status of the conceptualist position, by showing that the number of sentences lacking any psychological locus is unimaginably vast and that this collection includes sentences, equal in size to *every* transfinite cardinal. For such sentences, the notion of an actual psychological locus, even under the loosest of idealizations, makes no sense.

On the other hand, recognition of a realm of sentences equinumerous with the realm of sets raises absolutely no ontological problems not already implicit in standard set theory, problems which have to be faced by any viable ontological position. We conclude, therefore, that the demonstration that NLs are megacollections lends credibility to the realist position by showing, in another domain, the apparently insuperable problems facing any attempt to identify objects in the domain with aspects of the physical or psychological universe.

REFERENCES

Bartsch, R., and Vennemann, T., (1972) *Semantic Structures*, Athenaum Verlag, Frankfurt, W. Germany.
Brame, M. (1979) *Essays Toward Realistic Syntax*, Noit Amrofer, Seattle, Wash.
Bresnan, J. (1982) *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge, Mass.
Chafe, W. (1970*a*) *A Semantically Based Sketch of Onondaga*, Indiana University Publications in Anthropology and Linguistics, Vol. 36, No. 2, P. II, Bloomington, Ind.
— (1970*b*) *Meaning and the Structure of Language*. The University of Chicago Press, Chicago.
Chomsky, N. (1957) *Syntactic Structures* Mouton & Co., The Hague, Holland.
— (1965) *Aspects of the Theory of Syntax*, MIT Press, Cambridge, Mass.
— (1977*a*) *Reflections on Language*, Pantheon Books, New York.
— (1977*b*) *Essays on Form and Interpretation*, North-Holland Publishing Co., Amsterdam, Holland.
— (1981) *Lectures on Government and Binding*, Foris Publications, Dordrecht, Holland.
— (1982) *Some Concepts and Consequences of the Theory of Government and Binding*, MIT Press, Cambridge, Mass.
Dik, S. (1968) *Co-ordination*, North-Holland Publishing Co., Amsterdam, Holland.
— (1978) *Functional Grammar*, North-Holland Publishing Co., Amsterdam, Holland.
— (1980) *Studies in Functional Grammar*, Academic Press, New York.
Dowty, D. (1978) 'Governed Transformations as Lexical Rules in a Montague Grammar', *Linguistic Inquiry*, 9, pp. 393–426.
— (1982) 'Grammatical Relations and Montague Grammar', in P. Jacobson and G. Pullum (eds.), *The Nature of Syntactic Representation*, Reidel & Co., Dordrecht, Holland.
Eisenberg, M. (1971) *Axiomatic Theory of Sets and Classes*, Holt, Rinehart & Winston, New York.
Gazdar, G. (1981) 'Unbounded Dependencies and Coordinate Structure', *Linguistic Inquiry*, 12, pp. 155–84.
— (1982) 'Phrase Structure Grammar', in P. Jacobson and G. Pullum (eds.), *The Nature of Syntactic Representation*, Reidel & Co., Dordrecht, Holland.
Harman, G. (1963) 'Generative Grammars Without Transformational Rules: A Defense of Phrase Structure', *Language*, 39, pp. 597–616.
— (1982) *On Noam Chomsky*, 2nd edn., University of Massachusetts Press, Amherst, Mass.
Harris, Z. (1968) *Mathematical Structures of Language*, Interscience Publishers, John Wiley & Sons, New York.
Hays, D. (1964) 'Dependency Theory: A Formalism and Some Observations', *Language*, 40, pp. 511–25.
Hockett, C. (1966) *Language, Mathematics and Linguistics*, Mouton & Co., The Hague, Holland.
— (1968) *The State of the Art*, Mouton & Co., The Hague, Holland.
Hudson, R. (1971) *English Complex Sentences*, North-Holland Publishing Co., Amsterdam, Holland.

— (1976) *Arguments for a Non-transformational Grammar*, The University of Chicago Press, Chicago.

Johnson, D. (1979) *Toward a Theory of Relationally-Based Grammar*, Garland Publishing Inc., New York.

— and Postal P. (1980) *Arc Pair Grammar*, Princeton University Press, Princeton, NJ.

Joshi, A., Kosaraju, S., and Yamada, H. (1972) 'String Adjunct Grammars: I', *Information and Control*, 21, pp. 93–116.

Kac, M. (1980) 'Corepresentational Grammar', in E. Moravcsik and J. Wirth (eds.), *Syntax and Semantics, Volume 13: Current Approaches to Syntax*, Academic Press, New York.

Keenan, E. (1980*a*) 'Passive is Phrasal (Not Sentential or Lexical)' in T. Hoekstra, H. Van derHulst and M. Moortgat (eds.), *Lexical Grammar*, Foris Publications, Dordrecht, Holland.

— (1980*b*) 'A Conception of Core Grammar', unpublished manuscript.

Koster, J. (1981) 'Configurational Grammar', in R. May and J. Koster (eds.), *Levels of Syntactic Representation*, Foris Publications, Dordrecht, Holland.

Lakoff, G. and Thompson, H. (1975) 'Introducing Cognitive Grammar', in C. Cogen *et al.* (eds.), *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, Berkeley Linguistics Society, Berkeley, Calif.

Lamb, S. (1966) *Outline of Stratificational Grammar*, Georgetown University Press, Washington, DC.

Lambek, J. (1961) 'On the Calculus of Syntactic Types', in R. Jakobson (ed.), *Structure of Language and Its Mathematical Aspects*, American Mathematical Society, Providence, RI.

Langendoen, T. (1982) 'The Grammatical Analysis of Texts', in S. Allen (ed.), *Proceedings of the Nobel Symposium on Text Processing*, Almqvist and Wiksell, Stockholm, Sweden.

— and Postal, P. (1984) *The Vastness of Natural Languages*, Basil Blackwell, Oxford, England.

Lockwood, D. (1972) *Introduction to Stratificational Linguistics*, Harcourt Brace Jovanovich, Inc., New York.

Longacre, R. (1964) *Grammar Discovery Procedures*, Mouton & Co., The Hague, Holland.

Melčuk, I. (1981) 'Meaning-Text Models: a Recent Trend in Soviet Linguistics', *Annual Review of Anthropology: 1981*, 10, pp. 27–62.

Partee, B. (1975) 'Montague Grammar and Transformational Grammar', *Linguistic Inquiry*, 6, pp. 203–300.

— (1976) *Montague Grammar*, Academic Press, New York.

Postal, P. (1982) 'Some Arc Pair Grammar Descriptions', in P. Jacobson and G. Pullum (eds.), *The Nature of Syntactic Representation*, Reidel & Co., Dordrecht, Holland.

Reich, P. (1969) 'The Finiteness of Natural Language', *Language*, 45, pp. 831–43.

Sanders, G. (1972) *Equational Grammar*, Mouton & Co., The Hague, Holland.

Schachter, P. (1978) Review of R. Hudson, *Arguments for a Transformational Grammar*, *Language*, 54, pp. 348–76.

Woods, W. (1970) 'Transition Network Grammars for Natural Language Analysis', *Communications of the ACM*, 13, pp. 10–38.