On Some Continuous Properties in Language

T.G. BEVER and J.M. CARROLL

Columbia University and T.J. Watson Research Center I B M

1. INTRODUCTION

Many systematic phenomena in language are reflected in behavioral continua, rather than in discrete categories (Labov, 1972; Ross, 1974; Sankoff, 1978; Lakoff, 1973). This essay is concerned with the impact of such systematic statistical phenomena on the form of linguistic description. If linguistic grammars are allowed to exploit the descriptive power of non-categorical rules, more candidate grammars will be available for each set of facts: such proliferation of candidate descriptions reduces the empirical import of universal linguistic theory. Furthermore, if human language is defined in terms of a set of statistical principles (Greenberg, 1963), then the definition of "possible language" cannot be sharply drawn; rather, we can only describe a universally "likely" language.

In the present discussion we explore two examples of probabilistic properties of language, one a phenomenon of English and the other true of all languages. In each case we demonstrate that the phenomenon can be explained as the result of the interaction of categorical linguistic structures with a particular perceptual process, the processing of propositional relations. We also show that our explanation leads to the prediction of new facts.

Our considerations bear most directly on the nature of linguistic intuitions. We adopt the position that an acceptability intuition is the percept that results from an interaction of structural knowledge, behavioral processes, and the mechanisms of introspection. This interactive theory is distinct from the position that all reliable acceptability facts are pertinent to a "structural" linguistic description; and it is distinct from the position that all reliable facts are pertinent only to a processing description. Rather we seek to clarify the nature of language by distinguishing its impact on our intuitions from that of the behavioral systems that deploy it. In the course of our investigations we hope also to use linguistic and psycholinguistic theory as tools to clarify the nature of all introspective processes.

1.1 The Nouniness Squish in English. Ross (1974) noticed that a large number of syntactic processes differ in their acceptability along a continuum of nominal structures, ranging from the lexical noun to the complex clause. Some of these processes yield highest acceptability with nouns, some highest with clauses. In each case, Ross showed that there is an acceptability continuum for nominal structures intermediate in complexity between nouns and clauses, e.g., degraded or tenseless clauses, nouns that are deverbal nominalizations (but cf. Gazdar and Klein, 1978).

The syntactic process of plural agreement, for example, applies best to "noun-like" arguments. Conjunctions of very noun-like initial sequences like *winning and losing*, as in (1e), are perfectly acceptable as compound subjects of plural verbs.

(1a) *That he will win and that you will lose
(1b) *For him to win and for you to lose
(1c) ?*Him winning and you losing   are inevitable.
(1d) ??His winning and your losing
(1e) (The) winning and (the) losing

More clause-like initial sequences, such as Poss-ing nominals in (1d), elicit increasing unacceptability, as indicated by Ross's ??-?*-* notational system.

The converse situation is illustrated by the process of extra-position. Very clause-like sequences like the sentential subject *that Max whistles* (2a) are perfectly acceptable as extraposed sentence-final sequences signaled by the dummy subject *it*.

(2) It appears to disturb Herbie
(a) that Max whistles.
(b) for Max to whistle.
(c) ?Max whistling.
(d) ?*Max's whistling.
(e) *(the) whistling.

However, more noun-like structures like the nominalization *Max whistling* (2c), are increasingly unacceptable in these frames.

Ross (1974) interpreted such data as proving the existence of structurally based "squishes" — in this case a squish from "noun" to "clause" with intermediate levels of nouniness/clausiness. Such empirical acceptability gradations between the extremes of noun-ness and clause-ness led Ross to claim that the categorial notions of "noun" and "clause" in linguistic theory must be given up in favor of a continuous gradient.

1.2 Upstairs Primacy Across Languages.   Ross (1972) noted that optional reordering rules apply more freely within main clauses than within subordinate clauses. For example (3b) is an optional version of (3a), resulting from an adverb fronting process that applies to verb-phrase adverbs. (Note that the same acceptability phenomena are not true of "sentence" adverbs, e.g. "fortunately, certainly. . . .", at least because they are positioned by different rules)
(3a) John left quickly
(3b) Quickly John left.
(3c) After John left quickly, Mary became upset
(3d) *After quickly John left, Mary became upset
The same process cannot apply in (3c), as evidenced by the unacceptability of (3d). Ross also noted a difference in the strength of the main/subordinate asymmetry as a function of the surface order of the clauses — the effects are much weaker in final position. For example, (3e) is more acceptable than (3d).
(3e) ?Mary became upset after quickly John left.

Ross observed that such restrictions are generally true across languages, although some languages may exhibit isolated exceptions. The claim is that *if* there is an asymmetry in the range of application of a given rule it is in favor of broader application to main clauses than to subordinate clauses.

It is not immediately clear how to state such facts as part of universal grammar. Since *some* syntactic rules may be required specifically for subordinate clauses, we cannot simply require that all syntactic rules at least apply to main clauses. The difference in *relative* acceptability of sentence-initial and sentence-final subordinate clauses with such rules compounds this difficulty.

1.3 Explaining Statistical Regularities.   The goal of studying linguistic structure is in part to isolate universals of language. Such universals are candidate universal properties of linguistic structure. Of course, certain universals, true of all known languages are not properties of universal grammar, but rather are due to non-

linguistic properties of human beings. For example, it is a universal of human languages that no language has a total lack of surface marking of deep structure relations. It is clear why such a property exists, namely to make the sentences in the language usable.

We can furthermore understand how the restriction would be naturally imposed on all existing human languages: children attempting to learn a language simply fail to master the use of linguistic rules or combinations of rules that would lead to intractable ambiguity (Bever 1970). However, to capture the relevant generalization as part of universal grammar would be unwieldy. The reason for this is that the notion of intractable ambiguity must be defined across the range of structurally possible devices which can signal deep structure relations, and that it refers to the sentence forms that are actually used in the language — not just to the *possible* sentence structures. (Bever and Langendoen, 1973; Carroll, 1980).

The same sorts of considerations arise in the description of language specific grammars. The acceptability of certain types of constructions, and the principles that these acceptability facts implicate, appear to be fundamentally behavioral — and not grammatical. The classic example is the unacceptability of multiply center embedded relative clauses (Miller and Chomsky, 1963) — which has defied a purely structural explanation (Bever, 1970; Carroll and Bever, 1976).

Clearly it is a theoretical advantage in interpreting the ontological and psychological status of grammar if its principles are categorical in nature, univocally governing the form of the rules and their procedures of application. However, there are many cases in which relative phenomena do not have behavioral explanations as obvious as those of intractable ambiguity and center embedding. In the following discussion we show that the primacy and nouniness phenomena are explicable as a function of experimentally and theoretically established principles of speech perception. Indeed, the same aspect of speech perception accounts for both phenomena, thus rendering them the result of an interaction of human speech behavior and linguistic structure.

2. SPEECH PERCEPTION

The listener is presented with an acoustic stimulus which can be analysed at a number of different levels, e.g. syllables, words, phrases, and so on. Sentence comprehension involves the perception of an integrated description of these different levels. The listener plays a relatively active role, formulating hypotheses about each level as he hears the sentence (Bever, 1975; Bever, Garrett, and Hurtig, 1973; Marslen-Wilson, 1973, 1975).

2.1 Segmentation Processes.   The processes of perception "segment" a heard sentence, recoding each segmented unit into an inner form. The teleology of segmentation processes is twofold: first, the listener must have some means of dividing sentences — which can be arbitrarily lengthly and/or complex — into smaller more manageable chunks (Miller, 1956); and second, by dividing the sentence into logical constituents, the listener can facilitate the perceptual reconstruction of the sentence into units that typically serve the goal of comprehension.

There are several typical behavioral indications of segmentation processing. Heightened processing load just prior to the ends of segmentation units is reflected in a decreased ability to carry out on-line ancillary tasks presented at that point (Abrams and Bever, 1969; Bever and Hurtig, 1975). Interruptions (e.g. clicks, tones, coughs) occurring during a sentence are subjectively heard at points between

segmentation units (Fodor and Bever, 1965; Bever, Kirk and Lackner, 1969; Chapin, Smith & Abramson, 1972; Carroll and Tanenhaus, 1978; Warren and Warren, 1976).

Material that has been segmented is recoded into a more abstract representation. This is evidenced by an abrupt drop in verbatim recall and word recognition performance after segmentation boundaries (Caplan, 1972; Sachs, 1967; Carroll, 1978). Recoding also releases processing capacity, as indicated by an increased ability to perform memory tasks interpolated immediately after the point of segmentation (Wanner and Maratsos, 1978; Carroll, 1979a).

2.2 Segmentation Units. Several structural hypotheses have been advanced regarding the specification of units of sentence comprehension. Initially, it was suggested that surface phrase structure defined the relevant units (e.g., Fodor and Bever, 1965). Subsequently, it was shown that surface constituents are effective segmentation units just to the extent that they reflect deep structure clause organization (Bever, Lackner, and Kirk, 1969). This led to the hypothesis that deep clausal structure defines the relevant units of sentence processing — and that one of the fundamental goals in sentence perception is to isolate logically complete and coherent propositional sequences (e.g., subject - verb - object).

However, the surface realization of a deep clause is frequently degraded — often to a single word. In (4) the gerund *singing* corresponds to the deep structure clause GOALIE SING.

(4) Singing was well known by all of the players on the local football squad to be the goalie's one passion in life.

However, isolating the word *singing* would not enable the listener to properly organize the propositional sequence to which it belongs — until 17 words later he encounters the subject noun *goalie*. Earlier we noted that processing load was one of the fundamental motivations for the construct of segmentation processing in the theory of sentence perception. But clearly a single word is an unlikely source of much processing load.

Accordingly, recent research has addressed the question of what cues in the surface sequence might inform the listener that a potentially propositional sequence may be isolated and recoded (Carroll, Tanenhaus, and Bever, 1978). Thus, instead of assuming that sentence segmentation units will find comprehensive definition at a particular level of syntactic structure, investigators have sought to define the "functional clause" as a compromise between the propositional structure of sentences and the perceptual processes which must reconstruct this structure during perception.

2.3 Functional Clauses. Functional clauses are optimally propositional, allowing listeners to recode them, and they are optimally long/complex, so as to utilize but not exceed the listener's segmentation processing capacity. Either or both of these properties may be compromised when necessary. A set of recoding strategies have been identified which are vehicles for this processing activity. The most well-studied of these recoding strategies is to listen for N-V-N sequences (Bever, 1970; Carroll, 1978). Given a noun-verb-noun sequence, the listener considers the possibility that some kind of a propositional subject-verb-object sequence has been isolated.

Carroll and Tanenhaus (1978) showed that brief tones interrupting sentences tended to be misheard as having occurred at or near a clause boundary more frequently when the preceding clause had the structure N-V-N. Carroll (1978) showed that immediate cued verbatim recall was slower if the cue originated in a prior N-V-N unit than the target, relative to cases in which no complete N-V-N unit separated the cue and the target. Carroll(1979a) found that performance on a memory task interpolated at the clause boundary was better when the initial clause

was N-V-N than when it lacked either the subject or the object noun.

A second strategy is to isolate verb inflections (Carroll, 1976, 1978). In general, sequences with a tense-bearing verb stem also carry a complete set of propositional relations. "Deverbal" stems (e.g., gerunds) will often be embedded in propositionally incomplete sequences (as indeed was the case in example (4) above). But the verb inflection strategy operates even when it is not redundant with the N-V-N strategy. In the immediate cued recall task, response times were slower when the cue occurred in a tensed N-V-N sequence than when the cue originated in an N-V-N sequence that was not tensed (Carroll, 1978).

A third strategy involves subordinate marking: Clauses that are marked as subordinate must be integrated as modifiers of independent or main clauses. Carroll and Tanenhaus (1978) found that sentence initial main clauses attracted tones to the clause boundary region more than did sentence initial subordinate clauses. Tanenhaus and Carroll (1975) found greater release of processing capacity after main than subordinate clauses. Townsend and Bever (1978) found that sentence initial main clauses are semantically recoded more quickly than subordinate clauses: immediate meaning recognition was better for main clauses, but immediate verbatim recall was poorer for main clauses compared with subordinate clauses.

Townsend and Bever found that subordination effects differ according to how semantically "causal" and structurally "independent" a subordinate clause is from its main clause. Subordinate clauses introduced by *if* or *because* are behaviorally similar to main clauses, while a subordinate clause introduced by *although* shows the most different effects from main clauses. Townsend and Bever argued that the internal content of a causal *if*-clause can be semantically analysed without reference to the main clause, while this is specifically not true of an adversative *although*-clause. This coordinates with the fact that an *if*-clause sets up an explicit "cause-effect" sequence relation between the two clauses, while an *although*-clause specifically *denies* such a relation with respect to some part of the *although*-clause.

A non-structural influence on segmentation strategies is sequence length. In the absence of any other information, the mere length of a sequence provides some incentive for the listener to initiate segmentation. Carroll (1976) found latency differences in the immediate cued recall task between three word long and seven word long sentence-initial noun phrases. Response latencies were greater for longer noun phrases, indicating that they were treated perceptually as units more than were the shorter noun phrases.

In brief, a substantial body of experimental research supports the claim that there is a behaviorally-based set of considerations which govern the functional "goodness" of a sequence as a carrier of a clause-like propositional unit.

3. THE BEHAVIORAL BASIS OF NOUNINESS AND PRIMACY.

In this section we explain the two linguistic phenomena, the nouniness squish and primacy. Our account rests on the theory and research about the isolation of the major perceptual units in sentence perception. It is obvious that the processes of perception are involved to some extent in rendering acceptability intuitions, since a sentence must be apprehended in some sense in order to be adjudged acceptable, ambiguous, etc. We can expect there to be cases in which the perceptual processes themselves modify our apparent intuitions.

3.1 Primacy Explained Since the theory of speech perception has an independent status in the theory of language, this approach is efficient: it rests on established concepts. For example, main/subordinate clause processing differences provide an explanation of the Primacy principle. Since subordinate clauses are retained in a

more literal form, departures from canonical surface order would increase processing load more than in the case of main clauses. This asymmetry in processing load does not predict that there always will be fewer optional movement rules that can apply to subordinate clauses in a language. But it does predict that if there is an asymmetry in the range of application it will be in that direction. That prediction is exactly the Primacy principle (see Bever, 1975).

The behavioral explanation also accounts for why the principle holds more strongly for initial than for final clauses. The processing difference between main and subordinate clauses should matter less when the subordinate is sentence-final. Townsend and Bever (1978) found behavioral evidence that in final position the subordinate/main processing differences are attenuated in the predicted way.

3.2 Nouniness Explained. Recall now the nouniness acceptability squishes presented in (1) and (2) above. The various sequence types within these examples do appear to differ systematically with respect to the presence of verb tense and N-V-N sequences. All of the a, b, c, and d examples have N-V-N sequences. The a-cases are sentential subjects and have in addition tensed verbs. The b cases have infinitival verbs. The c and d examples have nominalized verbs, but the d cases in addition bear the subordinating element *'s*. Finally, the e cases consist of nominalizations without N-V-N sequences.

The most dramatic contrasts in Ross's data obtain between the a and e cases. The a cases, N-V-N and tensed, are far more potentially effective segmentation units than the e cases, which are neither N-V-N nor tensed. Between these extremes the sequences order themselves in terms of how deverbal the verb element of the sequence is: the infinitives of the b cases are more verbal than the nominalizations, and the nominalizations with the possessive inflection are even less verbal (cf. nouns are objects of possession) (See Tenenhaus and Carroll, 1975).

What seems to be going on is this: to the extent that the listener characteristically recodes a sequence as a coherent segmentation unit, that sequence will be more acceptable in syntactic environments that select clauses and less acceptable in syntactic environments that select nouns. Squish effects, thus, rest upon an internal confusion between the application of a process to a category (e.g., "S") and the typical behavioral reflex of that category (e.g., "perceptual clause").

4. NEW FACTS

The behavioral account of nouniness and primacy extends immediately to a range of other empirical domains — the two explanations in fact interact to provide an account of a set of complex and otherwise puzzling facts.

4.1 A New Nouniness Squish Fact The account of speech perception given in section 3 predicts new levels in the squish, ones not specifically analyzed by Ross, but rather predicted by the principle that coherent sentence processing units are bearers of propositional structures. The contrast in (b) is such a case.
(5a) *It bothers Herbie the whistling.
(5b) ??It bothers Herbie the whistling by Max.
Ross did not analyze such cases, but the speech perception theory predicts that the sequence *the whistling by Max* should be more coherent — and therefore more clause-like from the viewpoint of the process of extraposition — than the sequence *the whistling*. The reason is that the former sequence completely represents the proposition MAX WHISTLE, the latter fails to represent the subject relation *whistle*. This account follows the general account of the nouniness squish presented above.

4.2 The Primacy Squish The differences between subordinate clauses characterized by Townsend and Bever allow us to *predict* the possibility of an

acceptability ordering of Primacy principle facts. "Causal" subordinates, like *if*, should show the weakest effects, while "adversative" clauses, like *although*, show the strongest effects. The sentences below seem to bear this out. It appears to us that the sentences from (6a) to (6c) are ordered in level of relative acceptability (just as they are ordered in level of relative causality and independence of the subordinate clause).
(6a) If quickly John left, everyone will be relieved.
(6b) ?When quickly John left, everyone was relieved.
(6c) ??Although quickly John left, everyone was relieved.

The primacy squish does *not* predict that *if*-clauses are less restrictive for all rules than *although*-clauses. For example the primacy squish predicts a pronominal command squish in which initial pronouns are *less* acceptable in initial *because*-clauses than *although*-clauses. This follows from the interpretation that a pronoun in a main clause cannot precede the first expression of the nounphrase it refers to. The more "like" a main clause the subordinate clause is, the less acceptable an initial pronoun in that clause should be. That is, such pronouns in *if*-clauses should be *less* acceptable than in *although*-clauses, as in (7). (We thank D. Townsend for calling these cases to our attention).
(7a) ?If he arrived late, something must have happened to Harry
(7b) Although he arrived early, something must have happened to Harry

4.3 Complex Compounds. Another area to which we have extended the account of nouniness is word formation. It is well known that in some cases noun *phrases* may participate in noun compounding.
(8a) a doll that is a little girl ⟶ a (little girl) doll
(8b) a doll that is a girl with a bike ⟶ a (girl with a bike) doll
Complex compound examples, like those in (8), become far less acceptable when the noun phrase to be inserted as the left-most constitutent contains tensed verbs and N-V-N sequences.
(9a) a doll that is a girl playing guitar ⟶ a (girl playing guitar) doll
(9b) a doll that is a girl that says "Mama" ⟶ a (girl that says "Mama") doll

When otherwise suitable segmentation sequences occur within a lexical structure, they may stimulate segmentation and recoding erroneously. Consider the plight of the listener, or intuiter, for sentence (10).

(10) For your birthday I will bring you a nice girl playing guitar/doll.
Notice first that the compound is hard to recognize, *girl* and *doll* are separated by the verb phrase *playing guitar*, which provides some processing loan encouragment of segmentation. Second, *girl playing guitar* can be (mis-)recognized as an N-V-N clausal unit. Finally, this mis-recognition is locally consistent with the rest of the information in the sentence — that is, the material prior to the slash constitutes a reasonable sentence by itself.

When complex compounds have less complete left-constituents, like *little girl doll*, these problems are mitigated, as in (11).
(11) For your birthday I will bring you a nice little girl/doll.
In this case, while the local information is still consistent with a segmentation at the slash position, nothing intervenes between *girl* and *doll,* and the suitability of *little girl* as an independent segmentation unit is low. (See Carroll, 1979b, for further discussion and report of related experimental research.)

4.3 The Nouniness-Primacy Squish. The behavioral explanation of the nouniness and primacy phenomena encourage behavioral consideration of their interaction. To examine this possibility, we embedded various structure types from the noun squish into subordinate constructions at different points in the Primacy squish. Consider (12), for example.

(12a) If it bothers Herbie that Max is whistling, that's too bad.
(12b) Although it bothers Herbie that Max is whistling, that's too bad.
(12c) ?if it bothers Herbie Max's whistling, that's too bad.
(12d) ?*Although it bothers Herbie Max's whistling, that's too bad.
(12e) ??If it bothers Herbie the whistling, that's too bad.
(12f) *Although it bothers Herbie the whistling, that's too bad.
The relevant generalization about these facts is that the effect of extraposition on the noun constituents is mitigated in *if*-clauses relative to *although*-clauses. These facts are predicted by the greater causality and independence of *if*-clauses. They occasion more segmentation following the object nounphrase *Herbie* which, in turn, isolates the extraposed nounphrase as a relatively independent perceptual clause. Conversely the more subordinate *although*-clause occasions less segmentation, leaving the extra-posed nounphrase to be incorporated as part of the preceding clause.

We have also isolated complementary cases, cases in which the fully acceptable sequence embeds the more noun-like constituent.
(13a) ?If that Max whistles bothers Herbie, that's too bad.
(13b) *Although that Max whistles bothers Herbie, that's too bad.
(13c) If Max's whistling bothers Herbie, that's too bad.
(13d) Although Max's whistling bothers Herbie, that's too bad.
(13e) If the whistling bothers Herbie, that's too bad.
(13f) Although the whistling bothers Herbie, that's too bad.
In these cases as well the acceptability is lower in the *although*-clauses, and particularily so at the most clause-like end of the noun-clause continuum. The explanation of these facts rests on the relative complexity of a complex subject embedded within a subordinate clause. During processing, subordinate clauses are retained in a relatively sequential form. This predicts that a complex non-canonical internal structure will cause greater processing load, than in main clauses.

Also consistent with our analysis of the primacy phenomenon, the unacceptability of the various subordinate types is mitigated when the critical clauses are postpositioned, as in (14).
(14a) That's too bad if it bothers Herbie the whistling.
(14b) ?That's too bad although it bothers Herbie the whistling.
(14c) That's too bad if that Max whistles bothers Herbie.
(14d) ?That's too bad although that Max whistles bothers Herbie.

5. THE REAL NATURE OF INTUITIONS.

The most basic linguistic facts reside in the personal intuitions rendered freely by native speakers of a language. Such data provide a simple and direct vehicle for testing theories of linguistic representation. Nevertheless, intuitions are not empirical primitives but complex behavioral performances in their own right. Methodological studies of intuitions have documented a range of relevant but extragrammatical factors (Bever, 1970; Carroll, Bever, and Pollack, 1979; Gleitman and Gleitman, 1970; Spencer, 1973). The present essay also makes this point: in order to experience an acceptability intuition for a sentence, one characterically apprehends the sentence — and the processes that organize speech perception will contribute to the ultimate nature of the intuition.

There are three obvious options for dealing with the sort of data we have considered here. First, we may with Ross assume that non-discrete data directly imply non-discrete theories of grammar. This is not satisfactory: Non-discrete grammar offers no account of why the continua are the way they are. The correspondances between such grammatical analyses and the predictions of our behavioral

account would have to be viewed as mere coincidence. And moreover, no distinction at all could be drawn between the squishy intuitions we have been concerned with here and the ineluctable intuitions upon which linguistic theory relies. A second option is to treat *all* acceptability phenomena as behavioral and non-structural (Clark and Haviland, 1974). This alternative also is inadequate: It cannot explain the categorical (un)acceptability of examples at either end of a continuum.

The third alternative is that examples with intermediate acceptability rest on an internal confusion by the informant between the application of a linguistic proces to a category (e.g., "S") and to the typical behavioral reflex of that category (e.g. "perceptual clause"). This explanation explains a variety of acceptability facts as well as predicting hitherto unnoticed ones. However, it is important to specify what the general conditions are that will lead to an acceptability squish. We might expect such a confusion to arise because of the everyday behavioral importance of extracting propositions, even from degraded surface forms. A further source of difficulty may be that in English, "S" can dominate "N" as well as the reverse; this property of mutual dominability excludes any structural way of disentangling the two categories in intermediate cases.

An underlying theme of our proposal is that *all* grammatical properties are categorical and that all apparent departures from this have a general explanation in an interactionist framework — a totally discrete system of grammar interacting with behavioral processes.