On the Distinction between Model-Theoretic and Generative-Enumerative Syntactic Frameworks*

Geoffrey K. Pullum[1] and Barbara C. Scholz[2]

[1] Department of Linguistics, University of California, Santa Cruz, California 95064, USA
(pullum@ling.ucsc.edu)

[2] Department of Philosophy, San José State University, San José, California 95192, USA
(scholz@ling.ucsc.edu)

Abstract. Two kinds of framework for stating grammars of natural languages emerged during the 20th century. Here we call them generative-enumerative syntax (GES) and model-theoretic syntax (MTS). They are based on very different mathematics. GES developed in the 1950s out of Post's work on the syntactic side of logic. MTS arose somewhat later out of the semantic side of logic. We identify some distinguishing theoretical features of these frameworks, relating to cardinality of the set of expressions, size of individual expressions, and 'transderivational constraints'. We then turn to three kinds of linguistic phenomena: partial grammaticality, the syntactic properties of expression fragments, and the fact that the lexicon of any natural language is in constant flux, and conclude that MTS has some major advantages for linguistic description that have been overlooked. We briefly consider the issue of what natural languages in MTS terms, and touch on implications for parsing and acquisition.

1 Introduction

The second half of the 20th century saw the emergence of two quite different types of frameworks for theorizing about the syntax of natural languages. One sprang from the syntactic side of mathematical logic, the other from the semantic side. They are more different than has been recognized hitherto. This paper contrasts them and highlights some of their theoretical and empirical differences.

* Early versions of some of this material were presented to the Chicago Linguistic Society, the Ohio State University, and the Australasian Association for Philosophy. We thank John Goldsmith, Lloyd Humberstone, Phokion Kolaitis, Paul Postal, Frank Richter, Jim Rogers, Arnold Zwicky for advice, comments, and assistance. We particularly thank Patrick Davidson, Line Mikkelsen, Glyn Morrill, Chris Potts, and Michael Wescoat for comments on the first draft of this paper. It should not, of course, be assumed that any of these people agree with what we say here; some of them definitely do not.

P. de Groote, G. Morrill, C. Retoré (Eds.): LACL 2001, LNAI 2099, pp. 17–43, 2001.

We begin by briefly sketching the two types of framework. In section 2 we survey three theoretical differences between them. Section 3 then points out the implications of three relevant types of linguistic phenomena that have been hitherto very largely ignored. Section 4 briefly touches on the issue of what natural languages are. Section 5 summarizes and concludes.

We should note that we use the term EXPRESSION throughout with a twofold ambiguity. First, it refers to both types and tokens of items in natural languages like words, phrases, and clauses. Second, it is used both for linguistic items and for idealized representations of their structures. When necessary, we disambiguate.

1.1 Generative-Enumerative Syntax

The first framework type we consider was developed in the 1950s on the basis of Emil Post's work on formalizing logical inference rules. Post's concern was the properties of rule systems for mechanically deriving inferences from an initial axiomatic sentence variable (Post 1943). His formalization was general enough to provide a new characterization of the recursively enumerable (henceforth, r.e.) sets. Chomsky (1959) applied Post's work on inference systems to the description of sets of strings more generally, defining a grammar is a finite device that generates a set of strings or other structures ('generates' in the sense of Post 1944, synonymous with 'recursively enumerates').

Chomsky's work defined the first framework for grammatical description of the type that we will call GENERATIVE-ENUMERATIVE SYNTAX (henceforth GES). It laid the foundation for the syntactic study of programming languages, and launched the subfield of formal language theory within computer science. But his application of Post's work to natural languages (Chomsky 1957) had even greater influence. Within natural language syntax, GES has been overwhelmingly dominant since 1957. Frameworks of the GES type include:

- all the types of phrase structure grammar in the Chomsky Hierarchy;
- transformational grammar in almost all its varieties, including those that used 'filters' as an overlay on a GES base to rule out some of the generated structures – as in Perlmutter (1971), or the government-binding (GB) theory of the 1980s;
- generalized phrase structure grammar as presented in Gazdar et al. (1985), which despite some components framed in terms of constraints on trees (e.g. the feature instantiation principles), still clearly falls within the GES frameworks;
- all forms of categorial grammar, which generate sets by providing a general rule for composing primitives (categorized lexical items) to construct larger units of structure;
- tree-adjoining grammars (Joshi 1987), which generate structure through combinatory operations on a finite set of primitives (known as 'initial trees');
- Chomsky's post-1990 'minimalist' work (e.g. Chomsky (1995);

- the abstract syntax of Keenan and Stabler (1996), under which a set of structures is generated by closing a set of lexical items under certain combinatory operations;

and many more frameworks that we do not have space to list. Notice that not all GES frameworks are formalized; the recursive definitions of formation rules for wffs given in logic textbooks – 'if A is a wff then $\neg A$ is a wff', and so on – can be seen as informal GES descriptions.

It is the dominance of GES within linguistics that has led to dynamic and procedural metaphors being endemic within natural language syntax today: GES naturally engenders such metaphors, since the steps in the process of generating structures are so easily visualized as operations taking place in real time.

1.2 Model-Theoretic Syntax

The second approach we consider emerged some time later, from developments on the semantic rather than the syntactic side of logic. It applies model theory rather than proof theory to natural language syntax. The work in computer science that it parallels lies in descriptive complexity theory (Ebbinghaus and Flum 1999; Immerman 1998) rather than formal language theory, its mathematical concepts and methods being typically set-theoretic rather than combinatory.

The idea is to state a grammar as a finite set of axioms in a formal logic with a model-theoretic interpretation. We refer to these axioms as CONSTRAINTS. The models of the constraints are the expressions that are described by the grammar. An expression is well formed only if it is a model of the theory. We refer to this approach to grammar as MODEL-THEORETIC SYNTAX (MTS).

Notice (since it will be important below) that expressions, not sets of expressions, are the models for an MTS grammar: an individual expression either satisfies or does not satisfy a grammar. An MTS grammar does NOT recursively define a set of expressions; it merely states necessary conditions on the syntactic structure of individual expressions.

There have been few developments within linguistics so far that could be described as fully and explicitly representative of the MTS approach; compared to GES, it has had very little influence. But among the kinds of work we would cite as exhibiting the seeds of MTS are:

- the non-procedural recasting of transformational grammar in terms of conditions on tree sequences that can be found (albeit not very explicitly) in Lakoff (1971);
- the pioneering but mostly ignored work on formalizing relational grammar by Johnson and Postal (1980), a work that (especially in chapter 14) makes several key points we mention below;
- Gerald Gazdar's reformulation of generalized phrase structure grammar in terms of conditions satisfied or not satisfied by individual trees, presented in unpublished lectures at Nijmegen and Stanford during 1987; and

– the effort in Gazdar et al. (1988), particularly the Appendix (pp. 15–17), to describe finite labeled trees by directly imposing conditions on them using a modal logic; and
– head-driven phrase structure grammar in its more recent forms, as discussed by Sag and Wasow (1999, ch. 16), King (1999), and Pollard (1999).[1]

Some other frameworks might be added; for example, the construction grammar of Fillmore and Kay (1999), which is informally presented but appears to contain no GES elements, and perhaps recent lexical-functional grammar and some varieties of optimality-theoretic syntax (though the latter is by no means fully clear; we return to the matter briefly in section 3.1).

These separate lines of research in linguistics reveal gradual convergence on one idea: that grammars might be framed as sets of direct constraints on expression structure rather than devices for recursively enumerating sets of expressions. Nonetheless, MTS did not really begin to take explicit form until the 1990s, when the idea of natural language expression structures as models of grammars framed as statements in a logic really began to coalesce in the work of Marcus Kracht (1993, 2001), Patrick Blackburn and his colleagues (Blackburn, Gardent and Meyer-Viol 1993, Blackburn and Meyer-Viol 1997, Blackburn and Gardent 1995), and James Rogers (1996, 1997, 1998, 1999).

However, in a sense even the work of these researchers has been done in the shadow of GES. It has largely focused on comparing MTS and GES. Kracht (1993) attempts to clarify the content of GB; Blackburn and Gardent (1997) proposes a way to give an MTS account of the lexical functional grammar of Bresnan and Kaplan (1982); and Rogers develops an MTS characterization of the tree-sets that are generable by context-free grammars (henceforth, CFGs), using it to give an MTS restatement of the linguistic content of two theories couched in GES terms, mid-1980s GPSG (Rogers 1997) and mid-1980s GB (1998).

Given the enormous influence of GES frameworks, early advocates of MTS needed to show that sets of expressions could be defined, because in GES frameworks sets of expressions are identified with the object of study, i.e., the natural language being described. But the concentration on how MTS can simulate GES has led to certain differences between the two kinds of framework going unnoticed. The main aim of this paper is to consider MTS on its own terms, highlighting some of its overlooked features.

2 Theoretical Divergences

2.1 Cardinality of the Set of Expressions

Any grammar of the GES type generates a set, with a fixed, definite number of expressions as elements. No grammar of the MTS type entails that the grammatical expressions form a set of some definite cardinality.

[1] Note in particular Sag and Wasow's remark (1999: 382) that a grammar is "nothing more than a set of descriptions of typed objects . . . the constructs of the grammar are no longer clauses in a recursive definition of linguistic structures."

For GES the matter is purely definitional. A GES grammar states the characteristic function of a specific set of expressions. In the formal language theory literature the set in question is notated $L(G)$ for a grammar G.[2] To say that G generates a specific set means that for each element x of the set, G licenses a derivation, i.e. a finite set of steps that amounts to a proof that x is grammatical according to the rules of G.

Under certain conditions $L(G)$ will be denumerably infinite. For example, let $G = \langle V_N, V_T, S, P\rangle$ be an e-free CFG in which no nonterminals are useless (i.e. every nonterminal is accessible from S and has some positive yield). If for some $\alpha \in V_N$ we have $\alpha \stackrel{+}{\Longrightarrow} \psi_1\alpha\psi_2$ (where $\psi_1\psi_2 \neq e$), then $L(G)$ is infinite, and in all other cases $L(G)$ is finite.

The property of licensing derivations that allow a constituent labeled α as a proper subconstituent of another such constituent – which we will henceforth call the α-within-α property – is decidable for CFGs (though not, by an application of Rice's theorem, for context-sensitive grammars or URSs).[3]

If the conditions for generating an infinite language are not met, on the other hand, there will be just a finite number of expressions generated (and at least in the case of CFGs, the number will be computable from G).

Turning to MTS, we find some significant differences, though they are subtle and have been generally overlooked. MTS grammars are in effect statements in a logic, and although such statements can determine an upper bound on the number of their models, it takes a rather special kind of statement to do it. As an example, consider the MTS grammar $\mathit{\Gamma}_1$ over a node-label vocabulary $\{B, S\}$ and terminal vocabulary $\{a, b\}$, consisting of the union of the usual axioms for trees (for concreteness, those in Rogers 1998: 15–16)[4] conjoined with (1).

(1) $\exists x \exists y[S(x) \wedge a(y) \wedge x \triangleleft y \wedge \forall z[z \neq x \rightarrow z = y]]$

(where $S(x)$ means 'x is labeled by the nonterminal symbol S' and $a(y)$ means 'y is labeled by the terminal symbol a'). There is only one tree that models it, the tree with S-labeled root and one a-labeled daughter. But most MTS grammars do not determine a bound on the number of their models in this way. If we remove the constraint '$\forall z[z \neq x \rightarrow z = y]$' from $\mathit{\Gamma}_1$ we get a much more permissive grammar, namely the tree axioms plus (2):

(2) $\exists x \exists y[S(x) \wedge a(y) \wedge x \triangleleft y]$

This grammar, which we will call $\mathit{\Gamma}_2$, has arbitrarily many models, since it requires only that there be SOME S-labeled node with an a-labeled daughter.

[2] In what follows we adopt the convention of using G as a variable over GES grammars but $\mathit{\Gamma}$ for MTS grammars.

[3] To see this, consider the relation 'feeds' that holds between a rule $R_1 = \alpha \rightarrow \ldots \beta \ldots$ and a rule $R_2 = \beta \rightarrow \ldots$, for some $\alpha, \beta \in V_N$. We can construct a finite graph of the 'feeds' relation for any CFG, labeling the vertices with rules and adding an edge from the R_i vertex to the R_j vertex whenever R_i feeds R_j. G has the α-within-α property, and hence $L(G)$ is infinite, iff the graph of the 'feeds' relation contains a cycle, and this can be determined straightforwardly by inspection of the graph.

[4] Rogers' axiom $A4$ should read: $(\forall x, y)[x \triangleleft^+ y \leftrightarrow (x \triangleleft^* y \ \wedge \ x \not\approx y)]$.

But while a grammar like Γ_2 is COMPATIBLE with there being infinitely many expressions, it is also compatible with there being only some finite number of expressions, or even none. In all but the rather peculiar case of grammars like Γ_1, the issue of how many expressions there really are is a question to be settled independently of the grammar.

Some logicians may object at this point that there certainly is an answer to how many models Γ_2 has: it demonstrably has $\aleph_0$ models, since the set of finite labeled trees that satisfy it can be enumerated in lexicographical order and put into one-to-one correspondence with the natural numbers.

The response to this objection is as follows. We grant that if $\mathcal{T}$ is the set of all finite labeled trees over a given finite vocabulary of labels, then the set $\mathcal{T}_{\Gamma_2}$ defined by (3) is denumerably infinite.

(3) $\mathcal{T}_{\Gamma_2} \stackrel{\text{def}}{=} \{\tau : \tau \in \mathcal{T} \wedge \tau \models \Gamma_2\}$

However, notice that the cardinalities of the sets $\mathcal{T}$ and $\mathcal{T}_{\Gamma_2}$ are not determined by Γ_2. These sets are specified in the metatheory. Indeed, not only are they not defined by Γ_2, neither of them can be defined by ANY grammar written in the same metalanguage: Γ_2 is a first-order statement, and the property of finiteness is not first-order expressible (by an easy corollary of Trakhtenbrot's theorem for arbitrary first-order structures).

Natural language syntax must (at least) describe the shared structural properties of an indefinite number of distinct expressions. What is important is that there is no definite upper bound on how many expressions are described. GES satisfies this desideratum on grammars by generating a denumerably infinite set of expressions, and MTS satisfies it by NOT EXCLUDING the possibility of there being indefinitely many models for the grammar.

2.2 Finite and Infinite Expressions

We now turn to a point distinct from the previous one but intimately related to it: in a GES framework, a grammar always fixes a finite size for each grammatical expression, while for grammars in an MTS framework this is not the case.

Again, in GES frameworks this is a trivial consequence of the definition of a grammar. For an expression x to be in $L(G)$ for some GES grammar G, it must be possible to derive x in a finite number of steps using the rules of the grammar, and each rule either introduces at most a finite number of new symbols (in a top-down grammar) or combines a finite number of symbols (in a bottom-up grammar).

Again, with MTS things are quite different. Given an MTS grammar Γ with trees as models, there may be infinite objects satisfying Γ. In fact, if an MTS grammar is given in a first order language and there are finite trees of arbitrary size that are models of the grammar, there MUST be an infinite model – this is a corollary of compactness. The grammar Γ_2 considered above, for example, clearly has infinite models.

No absolute advantage or disadvantage accrues to MTS, or to GES, in virtue of this observation.[5] GES grammars can in principle be modified to permit generation of infinite strings with infinite parse trees (see Thomas 1990). And the issue of whether an MTS grammar can exclude the possibility of expressions of infinite size depends on the choice of the metalanguage for stating grammars: using a second-order logic for stating constraints it is possible to restrict models to the finite by statements in an MTS grammar. It is likewise possible using modal logic: **S4Grz** (which is in effect the logic used by Gazdar et al. 1988 to describe feature trees: Kracht 1989 proves the equivalence) is a modal logic that exactly characterizes the set of finite trees, as noted by Hughes and Cresswell (1984: 162–3, n. 5).

The point is not about absolute advantage but about tradeoffs in capability. GES grammars, by definition, generate r.e. sets and fix cardinalities for them. MTS grammars do not: defining a specific set containing all intended models of a certain sort for an MTS grammar must be done in the metatheory. It follows that GES in its classic form blocks the inclusion of any infinite expression in the generated set, whereas MTS by default allows infinite models to satisfy a grammar.

2.3 Optionality and Transderivationality

A much more important difference emerges when we focus on a key feature of MTS grammars, namely that their models are individual expression structures. This feature of MTS has implications for universal syntactic theory. No MTS grammar quantifies over whole expression structures. For example, if trees are the intended models, quantifiers in the statements of an MTS grammar range over a set of nodes, not over a set of trees. A rather striking consequence of this emerges when we re-examine the principles of X-bar theory.

We review the principles very informally first. X-bar theory takes the nonterminal vocabulary (those with which internal nodes are labeled) to be partitioned into several sets known as bar levels. 'X' is used as a variable over the atomic symbols or feature complexes needed for lexical category distinctions ('finite transitive verb', 'singular proper noun', etc.). Node labels can be represented in the form X^i, where for each $i \geq 0$ the category label X^i belongs to bar level i. (Intuitively, a phrase of bar level i is a phrase i levels of inclusiveness up from the lexical category constituent of X^0 that is its ultimate lexical head.) A maximum bar level m is a constant fixed by the specific theory. The full nonterminal vocabulary V_N is the union of all X^i for $0 \leq i \leq m$. A constituent labeled X^m for some X is referred to as a *maximal projection*, or simply as *maximal*. The principles of the theory can be stated very informally thus:

[5] The argument of Langendoen and Postal (1984), to the effect that GES grammars are falsified by their failure to describe infinite expressions, seems to us completely misguided.

(4) PRINCIPLES OF X-BAR THEORY

Lexicality:	for every X^i ($i \geq 0$) there is an X^0;
Uniformity:	for every X^0 there is an X^m;
Succession:	every node X^i node ($i \geq 1$) has an X^{i-1} head daughter;
Centrality:	the root node label is X^m for some X;
Maximality:	non-head daughters are maximal;
Optionality:	non-head daughters are optional.

Kornai and Pullum (1990) state these principles entirely in terms of a GES framework (specifically, in terms of statements about sets of CFG rules). However, it is actually impossible to make all of the above principles precise in an MTS framework. We can show this by contrasting Maximality with Optionality.

The apparent similarity between the statements 'non-heads are maximal' and 'non-heads are optional', using English adjectives, disguises a fundamental difference. Maximality can be defined on trees considered individually. Optionality cannot. The property of a node that makes it maximal is that its category label X^i is such that $i = m$. There is no property of a node, or relation between nodes, that makes it optional. Whether a constituent in a tree is optional depends on properties of other trees.

Specifically, optionality of a constituent can be defined only by quantifying over the trees belonging to some set. Let τ be a local subtree of some tree in a tree-set $\mathcal{T}$. Then a daughter α in τ is optional w.r.t. $\mathcal{T}$ iff erasing α from τ yields a local tree τ' that occurs in some tree in $\mathcal{T}$. Optionality is not a property of nodes in a tree at all, but a quantificationally complex property involving a different universe of discourse, namely a set of trees. Only by allowing quantification over trees can we express its content at all.

It should not be a surprise, then, that in the account of X-bar theory set forth in Rogers (1998) there is no mention of Optionality. This is not an oversight: Rogers' framework, like any MTS framework, simply does not permit it to be stated: his models are trees.

Whether this is an advantage or a major problem turns on whether Optionality is a sound principle of universal grammar. In actuality, it seems not to be. Though it is announced as a general principle in a number of works (for clear examples see e.g. Emonds 1976: 16, Base Restriction III, and Jackendoff 1977: 36, 43), it turns out to be incompatible with numerous uncontroversial facts about syntactic structure in natural languages (see Kornai and Pullum 1990: 34–35).

If this is correct – if it is a deep fact about natural languages that they do NOT respect the Optionality principle – then the absolute impossibility of expressing that principle in MTS terms may be seen as a virtue of MTS frameworks.

Related to this point is a more general one. Some GES theories in the 1970s posited TRANSDERIVATIONAL CONSTRAINTS, which excluded transformational derivations (i.e. finite tree sequences) on the basis of comparison with other derivations, e.g. by saying 'a structure with property π is well formed only if a structure $f(\pi)$ is well formed', or 'a structure with property π is well formed only if a structure $f(\pi)$ is not well formed', where f is a function from properties of derivations to properties of derivations.

For example, it has occasionally been claimed that certain natural languages permit constituent order inversions up to, but not beyond, the point where ambiguity would arise. Russian is a case in point. Its basic constituent order is Subject-Verb-Object (SVO), but stylistic inversion yielding Object-Verb-Subject (OVS) constituent order is often found. Hetzron (1972: 252–3), citing an earlier claim by Roman Jakobson, claims that in exactly those cases where the sequence of constituents $N_a\ V_b\ N_c$ in the clause has a corresponding SVO clause $N_c\ V_b\ N_a$ that is also grammatical, so that ambiguity arises between SVO and OVS interpretations of the same string, Russian syntax forbids inversion to OVS. For example, he claims that (5a) cannot be expressed in the form (5b).

(5) a. *mat'* *rodila* *doč*
mother gave-birth-to daughter
'The mother gave birth to the daughter.'
b. *doč* *rodila* *mat'*
daughter gave-birth-to mother

Intuitively this is because neither case-marking on the nouns *mat'* and *doč* nor gender-marking agreement on the verb *rodila* differentiates (5b) from the SVO clause that would have identical form.

Such a constraint is unstatable in an MTS grammar. Again this is because the models are individual expression structures. A question therefore arises about whether ambiguity-avoidance constraints are ever required in grammars. We think not. There are certainly functional considerations that (sometimes) militate against ambiguous use of a language by its speakers: people do sometimes attempt to make sure they choose expressions with forms that convey their intended meaning unambiguously. But it has repeatedly been found that building such functional pressures into syntactic constraints is a mistake. Both theoretical and empirical considerations suggest that grammars of natural languages do not ban ambiguity.

In the case of Hetzron's claims, one suspicious aspect of his analysis of Russian is that his description calls for a syntactic constraint to be sensitive to various aspects of phonological and phonetic realization of morphosyntactic feature (not just, e.g., whether a given noun is accusative, but whether its form contains a distinct accusative marker). There is a compelling weight of evidence against describing natural languages in such terms (see Miller et al. 1997 for references on the Principle of Phonology-Free Syntax, and discussion of some instructive examples). But there is also direct empirical evidence: Hetzron's (and Jakobson's) claims about Russian are just not true. OVS clauses in which neither NP case nor verb agreement disambiguate OVS from SVO are found in texts. There is syntactically wrong with Majakovskij's line in (6), where the allegedly forbidden ambiguous OVS appears twice.

(6) *Teplo daet peč, peč piaet ugol'*
warmth gives stove stove feeds coal
'The stove gives warmth, coal feeds the stove.'

Such examples appear in all kinds of text; Buttke (1969: 57) notes that up to 3% of cases of O before S in Russian prose are ambiguous in this way.[6]

If it is true that ambiguity-avoidance constraints do not figure in the syntactic constraints of the grammars of natural languages, then it may be seen as a positive virtue for a framework to be unable to express them.

3 Phenomena

In this section we argue that there are three related kinds of linguistic phenomena that are better described in MTS frameworks than in GES. In these domains at least, MTS frameworks have both broader scope and greater accuracy in describing the phenomena.

3.1 Gradient Ungrammaticality

Anyone who knows a natural language knows that some utterances are not completely well formed. Speakers produce utterances that even they would agree are grammatically imperfect – not by some external authority's standard but by their own. But experienced users of a language are also aware that some ungrammatical utterances are much closer to being grammatical than others.

We take this feature of utterances to be also a feature of expressions – or rather (since it may be better to limit the term 'expression' to what is fully grammatical) those objects that are like expressions except that they are only partially well-formed; let us call these QUASI-EXPRESSIONS. What we are saying is that some quasi-expressions are closer to being grammatical than others. In consequence, any framework for describing syntactic structure that can also describe degrees of ungrammaticality for quasi-expressions is to be preferred to one that cannot.

Some take a different view. For example, Schütze (1996) assumes that no GES grammar should generate any quasi-expression, of any degree of ungrammaticality. Rather, any perceived degree of ungrammaticality is a property to be described by performance processing mechanisms, not by the grammar. We believe this is the wrong view of partial ungrammaticality, and is probably an artifact of the difficulty GES frameworks have in describing that phenomenon.

MTS states grammars in a way that offers an elegant and heuristically suggestive starting point for thinking about the comparative ungrammaticality of quasi-expressions. Recall that an MTS grammar $\Gamma = \{\varphi_1, \ldots, \varphi_n\}$ is a finite set of statements in which each φ_i must be satisfied by any expression if the expression is to count as (fully) well formed. Γ may also be satisfied by various

[6] Thanks to Bernard Comrie for the Russian references.

quasi-expressions. Given any arbitrary set of expression structures and quasi-expression structures, we can use an MTS grammar Γ to define on that set a partial order that, intuitively, holds between two structures when one is, according to Γ and relative to the specified set, AT LEAST AS CLOSE TO BEING GRAMMATICAL as the other.

Let $\mathcal{U}$ be some universe of labeled trees, and let Γ be an MTS grammar with trees as models. Then there is a partial ordering $\trianglelefteq^{\mathcal{U}}_{\Gamma}$ defined on $\mathcal{U}$ by

$$(7) \quad \trianglelefteq^{\mathcal{U}}_{\Gamma} \stackrel{\text{def}}{=} \{\langle\tau_1,\tau_2\rangle : \tau_1,\tau_2 \in \mathcal{U} \ \wedge\ |\{\varphi : \tau_1 \models \varphi\}| \geq |\{\varphi : \tau_2 \models \varphi\}|\}$$

That is, $\tau_1 \trianglelefteq^{\mathcal{U}}_{\Gamma} \tau_2$ (for τ_1 and τ_2 in $\mathcal{U}$) iff τ_1 satisfies at least as many of the constraints of Γ as τ_2.

Clearly, a plausible starting point for describing degrees of ungrammaticality would be to assume that, other things being equal, a quasi-expression with structure τ_1 will be ungrammatical to a greater degree than a quasi-expression with structure τ_2 if and only if $\tau_1 \trianglelefteq^{\mathcal{U}}_{\Gamma} \tau_2$ for all suitable $\mathcal{U}$.

This is the most basic version, but various ways of refining the proposal immediately suggest themselves: the constraints of the grammar might fall into different classes determining the strength of their influence on ungrammaticality: constraints governing the inflection of words might be outranked by constraints applying to the composition of phrasal constituents in terms of the major categories of their daughters, and so on. (Ranking of constraints is, of course, a central idea of optimality theory, which we discuss further below.)

Notice that no extra machinery is called for: the suggested analysis of degrees of ungrammaticality simply exploits the content of the MTS grammar that is constructed to describe the fully grammatical expressions: on any set $\mathcal{U}$, there is a relation $\trianglelefteq^{\mathcal{U}}_{\Gamma}$ for MTS grammar Γ.

GES contrasts sharply in the extent of its failure to provide resources to define gradient ungrammaticality. A GES grammar G simply defines a set $L(G)$. It defines no ordering on the complement of that set, or for that matter on any set at all. The elements of $L(G)$ are perfectly grammatical, by definition, and where $\mathcal{Z}$ is any set, no element of $\mathcal{Z} - L(G)$ is described by a GES grammar for $L(G)$ as sharing any grammatical properties with any elements of $L(G)$. Arbitrary supplementary devices can be developed ad hoc to describe facts about degrees of ungrammaticality for non-members of $L(G)$, but the GES grammar G will contribute nothing to the project; all the work must be done by the extra machinery.

Three suggestions about accounting for gradient ungrammaticality deserve some brief discussion at this point. We will therefore digress to make a few remarks about (a) an early proposal of Chomsky's (1955, 1961), which provides a good illustration of what we mean about adding ad hoc machinery to a GES grammar; (b) the possible relevance of the optimality theory framework in this context; and (c) the implications of stochastic or probabilistic grammars.

Digression (a): Chomsky's Proposal. The failure of early GES grammars to represent degrees of ungrammaticality was noted quite soon after the publication of Chomsky (1957) by Archibald Hill (1961). Responding to the criticism, Chomsky (1961) published a proposed solution to the problem taken from a then unpublished work (Chomsky 1955), but the solution is not adequate. It involves, in effect, a function that maps members of the complement of the generated set to numbers representing the degree of their ungrammaticality. Given a language over a vocabulary V_T, Chomsky proposes defining a function $f : V_T^* - L(G) \mapsto \{1, 2, 3\}$ such that for any sequence $w \in V_T^* - L(G)$, the value of $f(w)$ gives a degree of ungrammaticality for w by comparing it with the expressions in $L(G)$ that it most closely matches in the sequence of lexical categories it instantiates.

The function f will be defined as follows. Let G be a transformational grammar with a lexicon that assigns the words (terminal symbols) to a set $K = \{\kappa_1, \ldots, \kappa_n\}$ of n lexical (bar-level 0) categories. Let λ be the binary relation holding between a word and a lexical (bar-level 0) category that it can realize according to the lexicon of G; that is, $\lambda(w_i, \kappa_j)$ for some $j \leq n$) means that word w_i is a permissible realization of lexical category κ_j. For a word sequence $w = w_1 \ldots w_m$ and a lexical category sequence $\kappa = \kappa_1 \ldots \kappa_m$, define $\lambda(w, \kappa)$ to mean $\forall i \leq m[\lambda(w_i, \kappa_i)]$. (From now on κ and κ' will be used as variables over lexical category sequences.) Then for a word sequence $w \notin L(G)$ and $i \in \{1, 2, 3\}$:

(8) $$f(w) = i \overset{\text{def}}{\iff} \exists\kappa\, \exists\kappa'\, \exists w'\ [w' \in L(G) \wedge \lambda(w, \kappa) \wedge \lambda(w', \kappa') \wedge \kappa \approx_i \kappa']$$

Crucial here are the relations $\approx_i$ of similarity between lexical category sequences, and these are as follows:

- $\approx_1$ is a relation of similarity among lexical category sequences that allows selection restriction features to be ignored (thus f(*Golf plays people*) $= 1$ because its lexical category sequence matches the sequence of a grammatical expression like *Golf amuses people* if we ignore the selection restrictions of *plays*: 'inanimate noun + transitive verb taking animate subject and inanimate object + animate noun' $\approx_1$ 'inanimate noun + transitive verb taking inanimate subject and animate object + animate noun');
- $\approx_2$ is a relation of similarity among lexical category sequences that allows strict subcategorization features to be ignored (thus f(*Golf elapses people*) $= 2$ because its lexical category sequence matches the sequence of a grammatical expression like *Golf amuses people* if we ignore the transitive subcategorization of *plays*: 'noun + transitive verb + noun' $\approx_2$ 'noun + intransitive verb + noun'); and
- $\approx_3$ is a relation of similarity among lexical category sequences that allows lexical category to be ignored altogether (i.e. it is the universal relation on K^*, relating every sequence to every other); thus f(*Elapses golf entertains*) $= 3$ because its lexical category sequence matches the sequence of a grammatical expression like *Golf amuses people* if we ignore the difference between nouns and verbs completely: 'noun + verb + noun' $\approx_3$ 'verb + noun + verb').

We see at least three difficulties with this proposal. The first is that it is not sufficiently fine-grained: there are in fact far more than three degrees of ungrammaticality. Consider:

(9) a. *I am the chair of my department.*
b. **I are the chair of my department.*
c. **Me are the chair of my department.*
d. **Me are the chair of me's department.*
e. **Me are chair the of me's department.*
f. **Me are chair the me's department of.*

Example (9a) is fully grammatical; (9b) is worse in virtue of one grammatical error; (9c) is worse than that (with two errors); (9d) is worse yet; (9e) is even worse; (9f) is worse still; and we could go on like this through arbitrarily many degrees of grammatical deviance. None of the degrees of difference in ungrammaticality in these examples is described under Chomsky's proposal (it does not cover deviance resulting from violations of constraints on inflection at all, as far as we can see).

The second difficulty is that the lack of any relationship between the proposed degrees of ungrammaticality and any specific violations of grammatical constraints leads to cases that intuitively involve multiple violations of constraints being represented as not differing in grammaticality from cases that intuitively involve only single constraint violations. For example, consider (10).

(10) **The car is in the the garage.*
D N V P D D N

This is just one incorrectly repeated determinative away from being fully grammatical.[7] Yet under Chomsky's proposal it is treated no differently from strings of complete gibberish like (11).

(10) **Of the and a but through or.*
P D Crd D Crd P Crd

Neither (10) nor (11) has a lexical category sequence matching that of any grammatical sentence of English, hence f maps both (10) and (11) to 3.

A third difficulty is that the whole proposal is nonconstructive and nonlocal. Instead of identifying the local sources of grammatical deviance (e.g., relating the ill-formedness of **They am here* to a failure of verb agreement), the proposal relates the status of a word sequence of a quasi-expression to the truth value of a statement that quantifies over the entire set of well-formed expressions. No algorithm for determining the status of an arbitrary word sequence is suggested, and for a transformational grammar there could not be one: because of the Turing-equivalence of transformational grammar, for a transformational grammar G, a word sequence $w \notin L(G)$, and a lexical category sequence κ such that $\lambda(w, \kappa)$, the question in (12) is undecidable:

[7] The category labels shown in (10) and (11) are as in Huddleston and Pullum (forthcoming): 'Crd' = Coordinator, 'D' = Determinative, 'N' = Noun, 'Nom' = Nominal, 'P' = Preposition, 'V' = Verb.

(12) $\exists w', \kappa' \, [w' \in L(G) \wedge \lambda(w', \kappa')]$?

Digression (b): Optimality-Theoretic Syntax. It might seem that optimality theory (OT) syntax would be highly relevant to the gradience of ungrammaticality: The central idea is that a grammar consists of a ranked (i.e. well-ordered) set of (putatively universal) constraints on structures called CANDIDATES relative to a given structure called the INPUT. (As we said above, it seems quite plausible to suppose that some constraints in natural language are of greater importance than others.) But in OT the constraint set is typically not consistent, so complete satisfaction of all the constraints is an impossibility. The expression structures defined as well formed are therefore not those that satisfy all the constraints. Whether a structure is defined as well formed is determined by a competition between candidate expressions regarding which does best at satisfying higher-ranked constraints.

To put it more precisely, an optimal candidate is chosen by finding the candidate α that in every pairwise competition with some other candidate β satisfies (or comes closer to satisfying, i.e. violates at fewer points) the highest-ranking constraint that distinguishes between α and β.

OT is distinguished, then, not by anything about the semantics of constraints, but by the semantics for a SET of constraints, which in OT is not the same as the semantics for the logical conjunction of those constraints.

The brief account given above of how degrees of ungrammaticality can be described by an MTS grammar may have seemed highly reminiscent of the way an OT grammar structures the set of candidates. Our view, however, is that although future research could vindicate such a view, it cannot be said to be borne out by the mainstream of the OT literature so far.

Much will turn on details of how OT grammars are interpreted. In most of the current literature a set of 'inputs' is assumed (what these are is highly obscure to us, but they may be something like representations of lexical items and semantically relevant grammatical relations including phrase membership). The grammar defines the optimal 'output' (some kind of more superficial representation of structure) for each, picking that optimal structure from a set defined by a (universal) function called `Gen`, which associates each input with a (generally infinite) set of candidates $\texttt{Gen}(I)$. The grammar defines the set of all pairs $\langle I, O \rangle$ such that I is an input and O is its optimal output – the candidate that wins when I is submitted to the algorithm that runs the pairwise competition with all the other candidates in $\texttt{Gen}(I)$.

In other words, the grammar determines a set just as a GES grammar does. Although in each competition there is a set of losing candidates, and that set is ordered (perhaps well-ordered) by the grammar, the order is not normally referred to in OT discussions of syntax, and for a given I no ordering is defined for any structures outside $\texttt{Gen}(I)$. Details of `Gen` are not normally supplied (OT researchers generally restrict attention to a handful of interesting alternatives for O for which membership in $\texttt{Gen}(I)$ is assumed to be obvious), so it is not possible to say how inclusive $\texttt{Gen}(I)$ is in a given case.

This is not a criticism of OT. We are merely pointing out that there is no simple relationship between OT and the MTS treatment of gradient ungrammaticality, since for the most part OT has not attempted to describe ungrammatical structures at all.[8]

Digression (c): Stochastic Grammars. Stochastic grammars have been the subject of increasing interest and attention within computational linguistics over the last decade. It has been suggested to us that they offer an effective defense of GES frameworks as regards the treatment of gradient ungrammaticality. We have pointed out that GES grammars simply define a set of perfectly well-formed expressions and do not say anything about degrees of approach to full grammaticality. Stochastic grammars differ from other GES grammars in that they associate probabilities with both rules and expressions, and describe some kinds of expression structures as having dramatically lower probability than others.

We see no basis for the suggestion that stochastic grammars have something to do with the topic of degrees of ungrammaticality, however. It seems to us that they do not address the issue at all. What a stochastic grammar does is to assign probabilities to the members of the set of generated expressions. For example, *You did it* can be represented as more probable than *I know you did it*, which in turn is more probable than *I know she knows you did it*, which in turn is more probable than *I know she knows I know you did it*, and so on down to vanishingly small (but still finite) probabilities. This is all very well – expressions with 37 degrees of subordinate clause embedding are indeed astronomically less likely to be encountered than those with 2 or 1 degrees of embedding. But (and this is the crucial point) an expression that is not generated by the grammar gets no probability assignment whatsoever.

Of course, a stochastic grammar that does not generate a given expression could always be revised to generate it and assign it a very low probability. But proceeding in that direction leads ultimately to a grammar that generates EVERY string over the terminal vocabulary; the probability assignments do all the work of defining what we commonly think of as well-formedness. This is a major departure from defining measures of relative probability for the grammatical expressions the grammar generates. It amounts to conflating the notion of ungrammaticality with the notion of extreme improbability. Some seem to have proposed such a conflation in the past (e.g. Hockett 1955: 10; Dixon 1963: 83–84), but the criticism by Chomsky (1957: 15–17) seems fully appropriate.

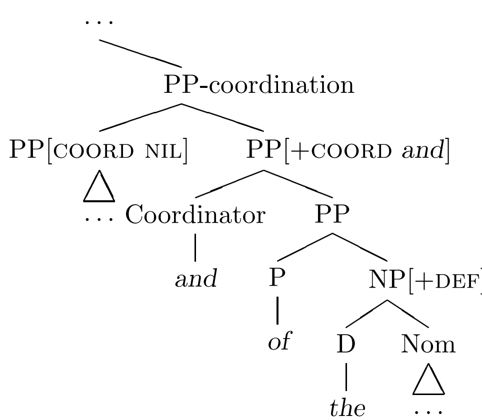
What MTS grammars can give an account of is the fact that quasi-expressions which everyone would agree are NOT grammatical nonetheless have syntactic structure. And this is something that stochastic grammars say absolutely nothing about.

[8] We note at least one exception: Keller (1998, 2000) has a very interesting treatment of the topic, in which selective re-ranking of constraints is used to shed light on degrees of well-formedness as assessed by acceptability judgments.

3.2 The Structure of Expression Fragments

Linguistic phenomena include not only complete expressions but also EXPRESSION FRAGMENTS. A fragment like *and of the* is certainly not generated by any GES grammar of English, and thus no GES grammar for English will describe its structure. It will be treated no differently from a random word string, like *the of and*, which does not share syntactic structure with any expression. Yet *and of the* has quite a bit of syntactic structure. We can represent part of it as in (13).

(13) THE STRUCTURE OF AN EXPRESSION FRAGMENT



The incomplete structure shown here is a syndetic coordination of prepositional phrases. The right coordinate daughter is marked [COORD *and*] (the '[CONJ NIL]' of Gazdar et al. 1985), and thus must have the Coordinator *and* as left branch. (The left coordinate daughter has no COORD value; either [COORD NIL] or [COORD *and*] would be grammatically possible.) The right sister of the Coordinator is a PP which has the preposition *of* as its head (left branch) and an NP object (right branch). This object NP is definite, and has as its left daughter a Determinative (D), namely *the*, the Determiner of the NP. The right branch is a Nominal (Nom), the head of the NP. The fragment would be seen in the context of full expressions like (14).

(14) *That cat is afraid of the dog and of the parrot.*

Our point is simply that while a GES grammar for English cannot describe the structure of any expression fragment, an MTS grammar will describe an expression fragment as satisfying all the relevant constraints on its structure. The absent left coordinate in the structure considered above satisfies every constraint of English (vacuously). Constraints like the one requiring a Preposition as head daughter of a PP, and the one requiring the Preposition *of* to have an NP object, are non-vacuously satisfied. The words that are present have their usual grammatical properties, and are combined as they ordinarily are. An MTS grammar does not require that an expression be complete in order to be described as having grammatical structure; the subparts of an expression have just the same structure as they do in fuller syntactic contexts.

The point just made is essentially the one made by Bresnan and Kaplan (1982: xlv–xlvii) under the heading of 'order-free composition', though there it is presented as a psychological claim. Bresnan and Kaplan point out (p. xlv) that "complete representations of local grammatical relations are effortlessly, fluently, and reliably constructed for arbitrary segments of sentences" by competent native speakers. We are pointing to a non-psychological analog of their claim: expression fragments have grammatical (and semantic) properties, but GES grammars do not describe them.

3.3 Lexical Flux

The third phenomenon we draw attention to is that the lexicons of natural languages are constantly in flux. No natural language has a fixed word stock. Each day and each hour new words are being added by creative recombination of combining forms (*Internet*, *Ebonics*, *futurology*), recategorizations (*Don't start Clintoning on me*), trade name coinages (*Celica*, *Camry*, *Elantra*), technical term creation (*bioinformatics*, *genomics*, *quark*), the invention of unprecedented personal names like *DeShonna* for girls (a familiar feature of African-American culture), spontaneous dubbings (*I will call him 'Mini-Me'!*), nonce uses (*Suppose we have a file called 'arglebargle'*), jocular coinages (*You bet your bippy!*), conventionalization of abbreviations (*MTV*, *IBM*, *CIA*), creation of acronyms (*NAFTA*, *AIDS*, *UNESCO*), onomatopoetic invention (*It went 'gadda-ga-DACK!' and stopped*), and so on.

Indeed, as noted by Harris (1968: 11), spoken expressions can even incorporate random vocal noises (of which there is arguably a nondenumerable infinity). Familiar utterances such as (15) suggest that the variety of material that can fill a categorized slot in a linguistic expression is in a sense not even bounded by the combinatory possibilities of some finite phonetic alphabet.

(15) *My car goes 'ehhrgh!' when I go over a bump.*

This is not an new observation. But it presents a serious difficulty for GES grammars, which enumerate sequences over a fixed terminal vocabulary. No sequence containing an element that is not in that fixed vocabulary is generated, and if a sequence is not generated, its structure is not described. It follows that the syntactic structure of any expression of English containing a novel lexical item goes undescribed by a GES grammar.

The point here is not that there would be some difficulty in modifying a GES grammar to make a new one that accommodated some novel lexical item. Our point is about the failure of a fixed GES grammar to describe phenomena correctly. Take a GES grammar for English over a vocabulary not including a word of the form *dibble* (for example). Such a grammar does not describe the structure of an expression like (16).

(16) *Do you have any of those little dibbles you had yesterday?*

Clearly it has the structure of an English expression despite the fact that *dibble* is not (as far as we know) in any dictionaries and has no meaning. The expression is a clause of closed interrogative type, with second person subject, plural direct object with an attributive adjective, bare relative clause containing a preterite verb, and so on and so on. Even *dibbles* has properties: it is inflected for plural number, it has a regular inflectional paradigm, and so on. A GES grammar over a vocabulary not including *dibble* does not generate the string at all and thus does not describe its syntax.

Bottom-up types of GES grammar, like categorial grammars or 'minimalism', exhibit this problem of not covering the phenomena in a particularly acute way: if *dibble* is not a lexical item, then it has no category; if it has no category, then it cannot be combined with *little* to make *little dibbles*; and that means there is no phrase of that form to combine with *you had yesterday* to make a phrase *little dibbles you had yesterday*; and so on. Absolutely no grammatical structure is built at all for an expression containing a nonexistent word.

The right claim to make about this word sequence is not that its structure cannot be described but that it is perfectly grammatical. Its oddness resides solely in the fact that since there are no lexical semantic constraints on *dibble*, the expression has a singularly large range of possible meanings. Novel lexical items do not by themselves make the syntactic structure of expressions containing them indescribable. Nor do they obliterate the structure of the expressions in which they occur.

This point is essentially the one made by Johnson and Postal (1980: 675–7), and they draw the same conclusion: MTS grammars make the right claims about expressions containing novel lexical items, and GES grammars make strikingly incorrect claims about them. And since the lexicon of any natural language is continuously in flux, the phenomena cannot be overlooked. Grammars should describe syntactic structure in a way that is not tied to the existence of any particular set of words in a lexicon. GES grammars conspicuously fail to do this.

4 Languages

So far in this paper we have not used the word 'language' without qualification. We use phrases like 'natural language', 'programming language', 'formal language theory', 'first-order language' and so on, but we avoid the use of the word 'language' in technical senses like those current in mathematical linguistics and GES frameworks, e.g. the sense 'set of strings'. In this section we argue that the use of the word 'language' for both the object of study in linguistics and the set defined by a GES grammar, has led to the mistaken view that a trivial consequence of GES frameworks is a deep discovery about natural languages. We propose a different view of what natural languages are.

4.1 The Supposed Infiniteness of Natural Languages

The claim that natural languages are infinite is not just dogma subscribed to by those who work within GES frameworks, but a shibboleth for the linguistics profession: failure to endorse it signals non-membership in the community of generative grammarians. Hardly any introductory text or course on generative linguistics fails to affirm it. The following three distinct claims are all made on the same page of Lasnik (2000: 3):

(i) "We need to find a way of representing structure that allows for infinity—in other words, that allows for a sentence inside a sentence inside a sentence, and so on."

(ii) "Infinity is one of the most fundamental properties of human languages, maybe the most fundamental one. People debate what the true linguistic universals are, but indisputably, infinity is central."

(iii) "Once we have some notion of structure, we are in a position to address the old question of the creative aspect of language use... The ability to produce and understand new sentences is intuitively related to the notion of infinity."

These claims are not elaborations of the same point; they are respectively about (i) the recursive structure of natural language syntax, (ii) the size of the sets generated by GES grammars, and (iii) the native speaker's ability to understand novel expressions.

The first claims that linguistic structure is sometimes recursive, and this is clearly correct: expressions as simple as *See Spot run* uncontroversially exhibit α-within-α structure – that is, constituents of some category α properly contained within larger constituents of category α. (*Run* is a verb phrase, and so is *see Spot run*, which properly contains it in *See Spot run*.) We fully accept that some natural language expressions exhibit α-within-α structure, of course.

Lasnik's second claim concerns something different: the existence of a denumerably infinite collection of expressions defined by a GES grammar. This is not a claim about the structure of expressions. It is supposed to be a claim about the size of natural languages. But in fact it is simply a trivial point about GES frameworks: basically, the set generated by a GES grammar G is countably infinite just in case G describes α-within-α structure. As we noted in section 2.1 (see footnote 3), this is a decidable property for some types of GES grammar, such as CFGs. And as we have just agreed, English has α-within-α structure: even a cursory examination of plausible structures for a few expressions of English reveals that VPs occur within VPs, NPs occur within NPs, and so on. Hence IF WE ASSUME THAT A NATURAL LANGUAGE CAN ONLY BE CORRECTLY DESCRIBED BY A GES GRAMMAR, it immediately follows that the set generated by a grammar for English contains infinitely many expressions.

But this is not a fact about natural languages. It is purely a fact about the properties of (a certain class of) GES grammars. Under an MTS framework it does NOT follow that the set generated by a grammar for English contains infinitely many expressions. From the fact that a grammar for a natural language should never specify that expressions have an upper bound on their length, and that some of their structures are α-within-α structures, no conclusion follows about how many expressions there are.

Notice that when applied to cultural products, inferences to infinity from lack of an upper bound clearly fail: there is no upper bound on the length of poems, but it does not follow that there are infinitely many poems. Similar reasoning also fails in biological domains: there is no set upper bound on length of lineage for an organism in generations, but it does not follow that there are, or could be, organisms with infinitely many descendants. To suppose an inference is valid in the case of expressions of a natural language is to confuse a property of GES grammars with a substantive claim about natural languages.

Turning now to the claim in (iii), we note that this is not a claim about either the structure of expressions or the size of a set of expressions, but about human linguistic capacities. To get the claim about capacities to follow from the claim about sets we need to assume not only that grammars generate sets but also that grammars describe speakers' linguistic capacities. We only get the conclusion that human linguistic capacities are infinite (in the sense that a speaker can in principle understand any of an infinite range of expressions) if we accept both. So a case for saying that human linguistic capacities are infinite must be made on the basis of two assumptions: (i) the assumption that GES grammars best describe languages, and (ii) the assumption that GES grammars best describe linguistic knowledge. Since the first assumption rests on a confusion, the second does too.

If instead we assume that natural languages are best described by MTS grammars, and that human linguistic capacities are too, we get the result that those capacities are creative and productive without any commitment to the existence of a determinate set containing all possible expressions. As Gareth Evans (1981) insightfully pointed out, creative and productive language use has nothing to do with infinity, and can even be realized (through the ability to recombine expression parts) within a finite language.

Moreover, other things are not equal in this case: since GES grammars do not cover the linguistic phenomena discussed in section 3, it is increasingly doubtful that natural languages are best described in GES terms.

4.2 Some Caveats and Clarifications

It should be stressed that we are NOT making the claim that an appropriate MTS grammar for natural language has only finitely many models. Given that any adequate grammar will describe α-within-α structure, the set $\mathbf{Mod}(\Gamma)$ of all models for an MTS grammar Γ is infinite. But that is a fact about what is true in the metatheory of the logic in which Γ is stated. By itself, Γ does not entail that there are infinitely many expressions, because an MTS grammar states only necessary conditions on the structure of expressions considered individually.

Note also that we do not deny that infinite sets of expressions are crucial in some areas of mathematical logic, computer science, formal learning theory, etc. Without reference to infinite sets of expressions in mathematically defined languages

- we cannot even talk about the compactness theorem for first-order logic in its usual form (if every finite subset of an infinite set of formulae has a model then the entire set has a model);
- we cannot construct complete theories about integer arithmetic because the truths about integers will outnumber the expressions of our language;
- computational complexity theory is trivialized because without an infinite set of algorithmic problem presentations in some language we can solve every problem in constant time by table look-up;
- formal learning theory is trivialized because the interesting problems of learning theory arise only with infinite classes of decidable languages (every class of finite languages is identifiable in the limit from text).

For these reasons and many others, the definition of infinite sets of expressions within various fields of logic and applied mathematics is essential.

But that is not the same as saying that it is appropriate for the framework linguists use in stating grammars to stipulate how many expressions there are in a natural language. Doubtless, it is sensible to ensure that the metalanguage in which we write MTS grammars for natural languages is NOT one in which every formula is guaranteed to have only finitely many models (see Thomas 1986, for example, where a class of logics with boundedly many variables is studied and shown to be of this sort). And if we adopt a grammar-statement metalanguage that allows a formula to have infinitely many models, it will probably be possible to define an infinite set $\mathcal{E}$ such that $\forall x[x \in \mathcal{E} \rightarrow x \models \Gamma_E]$, where Γ_E is a suitable MTS grammar for English. But English itself is not thereby equated with $\mathcal{E}$. To make that equation is to commit the error Lasnik commits concerning the alleged infiniteness of languages: it confuses a property of GES grammars with properties of natural languages and of human abilities.

An adequate MTS account of the syntax of a natural language will first and foremost accurately describe the structural properties of some large finite set of attested expressions. Uncontroversially, this will involve describing α-within-α structure. The 'creative aspect of language use' under this view is described by the way MTS grammars not only represent the structure of attested expressions correctly but also predict the structure of as yet unattested novel expressions, expression fragments, and quasi-expressions.

No statement about the size of the set of unattested expressions of a natural language plays any role in describing or explaining any linguistic data. Different views about the ontology of linguistics force different answers to how many expressions there are. Clearly, a thoroughgoing nominalist might accept that there are only finitely many expressions. And a platonist might claim that there are uncountably many of them (Langendoen and Postal 1984). A framework for syntax does not, and should not, decide such issues. Indeed, we have argued elsewhere that no framework or theory should rule out any purely ontological theory about what natural languages are (see Pullum and Scholz 1997 for some discussion of this point).

4.3 What Languages Are

We address one other issue, more tentatively, before concluding. It concerns the ordinary, common-sense notion of a language under which we can say that *The Times* in the UK, *The New York Times* in the USA, *The Sydney Morning Herald* in Australia, and other newspapers around the world, all publish in the same language – though of course we would not deny that there may be local differences concerning which expressions are judged grammatical by the relevant editors.

Generative grammar makes no attempt to reconstruct this notion. Indeed, Chomsky applauds the way GES frameworks have replaced the common-sense notion of a language with a stipulated technical concept: in ordinary parlance, he says,

> We speak of Chinese as "a language," although the various "Chinese dialects" are as diverse as Romance languages. We speak of Dutch and German as two separate languages, although some dialects of German are very close to dialects that we call "Dutch"... That any coherent account can be given of "language" in this sense is doubtful; surely none has been offered or even seriously attempted. Rather all scientific approaches have abandoned these elements of what is called "language" in common usage. (Chomsky, 1986: 15)

The idea seems to be that advances in theoretical linguistics may and should replace our ordinary concept of 'a language' with a theoretical one. Under GES, the commonsense notion of an expression being 'in English' is not conserved, and is not intended to be; it is replaced by the notion of belonging to a certain r.e. set. In ordinary parlance, the phrase 'in English' means something like 'structured in the English manner' ('in English' has manner adjunct function, and certainly does not mean 'in the r.e. set known as English'). The ordinary understanding of 'She spoke in English', for example, is fairly well captured by 'She spoke in the English manner.' An MTS grammar for English plausibly reconstructs this common-sense idea: the grammar is a set of statements that state conditions for being an expression that is structured in the English manner.

MTS frameworks therefore provide a better basis for a conservative reconstruction of the common-sense concept of a language, the one under which millions of different people may be correctly described as speakers of the same language. And contra GES advocates, we regard this as no bad thing. It may be that a select set of constraints can provide a useful description of the linguistic structure that is shared between the differing idiolects of the hundreds of millions of people around the world who can, in their different ways, be said to use Standard English. GES frameworks appear to have no chance of capturing such a notion.

5 Conclusions and Prospects

We have distinguished two types of framework for describing natural language syntax: GES, under which grammars enumerate sets of expressions, and MTS, under which grammars place necessary conditions on the structure of individual expressions. We noted three differences between them.

- GES grammars define sets of expressions with definition cardinalities, while MTS grammars do not.
- MTS grammars give structural descriptions that are (in some cases) satisfied by expressions of infinite size, while to do the same thing in GES terms would require redefining the fundamental notions of the framework such as 'derivation'.
- MTS frameworks, because their models are individual expression structures, are entirely unable to state any kind of generalization of the sort that was called 'transderivational' in 1970s generative grammar, or to express principles such as the putative Optionality principle of X-bar theory, while ways of doing both can be and have been explored within GES frameworks.

We have argued that there are well known linguistic phenomena that are more accurately and elegantly described by MTS. These same phenomena have been largely ignored by advocates of GES frameworks, perhaps because of the difficulty such frameworks have in describing them.

More questions at this point remain than have been answered. We need to explore which kind of logic will best describe natural languages:

- Are there PAROCHIAL syntactic properties that need second-order quantification, or is all such second-order quantification needed only in defining concepts of universal grammar?
- Is there a suitable description language that never commits to whether the class of all models is finite, denumerable, or nondenumerably infinite?
- Is there a suitable logic that guarantees that checking satisfaction of a grammar by an arbitrary model will be not just decidable but tractable?

And the familiar problems of processing and learning take on some new aspects. Given a solution to the segmentation problem of figuring out the sequence of words in an utterance as phonetically presented, the computational problem of processing can be informally stated, under MTS assumptions, like this:

(17) **The Parsing Problem:** Given a word sequence, find an expression structure that satisfies the grammar and has that word sequence as its frontier.

The point of parsing is to provide part of the basis for solving a much more difficult computational problem:

(18) **The Conveyed Meaning Problem:** Given an expression presented in some context of utterance, find the conveyed meaning of the expression in that context.

We need to know more about the complexity of these problems in terms that are not based entirely in GES assumptions about grammar. The 'dynamic syntax' of Kempson, Gabbay and Meyer-Viol (2001) is a welcome development in this regard, since its dynamic account of how meaning representations might be constructed in real time as a word sequence is presented word by word is in fact based on an MTS approach to syntactic description using a modal logic.

Ultimately, the hardest of all the problems in the language sciences awaits us: the problem of language acquisition. Learners of a natural language in effect accomplish, in MTS terms, something like a solution to the following computational problem:

(19) **The Acquisition Problem:** Given a sequence of expression structures paired with the contexts in which the expressions were uttered, devise a set of constraints on the structure and meaning of expressions that will permit success in parsing future unconsidered utterances.

This is an enormously hard problem: the primary linguistic data underdetermines every MTS grammar. But we close by pointing out one piece of good news that emerges from taking an MTS approach to syntax. The theorems presented by Gold (1967), which are often cited as proof that language learning from presented examples is impossible, will not be applicable. What Gold proves is often stated rather loosely, e.g. as that it is impossible to learn a grammar from nothing but exposure to a finite sequence of presentations of examples of grammatical expressions. More precisely, the result is this:

(20) *Theorem* (Gold): Let $\mathcal{L}$ be a class of string-sets containing all finite string-sets over a finite vocabulary V and at least some infinite string-sets over V. Then there is no algorithm that solves the following problem for all string-sets in $\mathcal{L}$: given a continuing sequence of presentations of elements taken from some member of $\mathcal{L}$, guess a GES grammar after each presentation, arriving after some finite number of presentations at a correct guess that is never subsequently changed.

What concerns us here is that Gold's results are entirely about algorithms for successful guessing of exact definitions of sets in the form of GES grammars (or equivalent recognition automata). The exact cardinalities of these sets are crucially relevant: a learning algorithm that is unable to determine from the presented data whether the right grammar generates an infinite set or not is defined as having failed.

Under the MTS view THERE IS NO SET DEFINITION TO BE GUESSED. The problem of devising a suitable constraint set on the basis of exposure to utterances in normal contexts of use remains apparently very hard, and the answer to whether the learner's cognitive system has innate linguistic priming is by no means settled (see Pullum and Scholz forthcoming for further discussion). But at least under MTS we do not start out under the cloud of a proof that the task is in principle impossible.

References

[1995]Backofen, Rolf; James Rogers; and K. Vijay-shanker (1995): A first-order axiomatization of the theory of finite trees. *Journal of Logic, Language, and Information* **4**: 5–39.

[1995]Blackburn, Patrick and Claire Gardent (1995): A specification language for lexical functional grammars. *Proceedings of the 7th EACL*, 39–44. European Association for Computational Linguistics.

[1993]Blackburn, Patrick, Claire Gardent, and Wilfried Meyer-Viol (1993): Talking about trees. *Proceedings of the 1993 Meeting of the European Chapter of the Association for Computational Linguistics*, 21–29.

[1997]Blackburn, Patrick and Wilfried Meyer-Viol (1997): Modal logic and model-theoretic syntax. In M. de Rijke (ed.), *Advances in Intensional Logic*, 29–60. Dordrecht: Kluwer Academic.

[1982]Bresnan, Joan and Ronald M. Kaplan (1982): Introduction: grammars as mental representations of language. In *The Mental Representation of Grammatical Relations*, ed. by Joan Bresnan, xvii–lii. Cambridge, MA: MIT Press.

[1969]Buttke, Kurt (1969): *Gesetzmässigkeiten der Wortfolge im Russischen.* Halle: Max Niemeyer.

[1955]Chomsky, Noam (1955): *The Logical Structure of Linguistic Theory.* Unpublished dittograph, MIT. Published in slightly revised form by Plenum, New York, 1975.

[1957]Chomsky, Noam (1957): *Syntactic Structures.* The Hague: Mouton.

[1959]Chomsky, Noam (1959): On certain formal properties of grammars. *Information and Control* **2** (1959) 137–167.

[1961]Chomsky, Noam (1961): Some methodological remarks on generative grammar. *Word* **17**, 219–239. Section 5 republished as 'Degrees of grammaticalness' in Fodor and Katz (1964), 384–389.

[1964]Chomsky, Noam (1964): *Current Issues in Linguistic Theory.* The Hague: Mouton. Page reference to reprinting in Fodor and Katz, eds.

[1986]Chomsky, Noam (1986): *Knowledge of Language: Its Origins, Nature, and Use.* New York: Praeger.

[1995]Chomsky, Noam (1995): *The Minimalist Program.* Cambridge, MA: MIT Press.

[1963]Dixon, Robert W. (1963): *Linguistic Science and Logic.* The Hague: Mouton.

[1999]Ebbinghaus, Heinz-Dieter, and Jorg Flum (1999): *Finite Model Theory.* Second edition. Berlin: Springer.

[1976]Emonds, Joseph E. (1976): *A Transformational Approach to English Syntax.* New York: Academic Press.

[1981]Evans, Gareth (1981): Reply: syntactic theory and tacit knowledge. In S. H. Holtzman and C. M. Leich (eds.), *Wittgenstein: To Follow a Rule*, 118–137. London: Routledge and Kegan Paul.

[1999]Fillmore, Charles W. and Paul Kay (1999) Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. *Language* **75**: 1-33.

[1964]Fodor, Jerry A. and Jerrold J. Katz, eds. (1964): *The Structure of Language: Readings in the Philosophy of Language.* Englewood Cliffs, NJ: Prentice-Hall.

[1985]Gazdar, Gerald; Ewan Klein; Geoffrey K. Pullum; and Ivan A. Sag (1985): *Generalized Phrase Structure Grammar.* Oxford: Basil Blackwell; Cambridge, MA: Harvard University Press.

[1988]Gazdar, Gerald; Geoffrey K. Pullum; Bob Carpenter; Ewan Klein; Thomas E. Hukari; and Robert D. Levine (1988): Category structures. *Computational Linguistics* **14**: 1–19.

[1967]Gold, E. Mark (1967): Language identification in the limit. *Information and Control* **10**: 441–474.

[1968]Harris, Zellig S. (1968): *Mathematical Structures of Language* (Interscience Tracts in Pure and Applied Mathematics, 21). New York: Interscience Publishers.

[1972]Hetzron, Robert (1972): Phonology in syntax. *Journal of Linguistics* **8**, 251–265.

[1961]Hill, Archibald A. (1961): Grammaticality. *Word* **17**, 1–10.

[1955]Hockett, Charles F. (1955): *A Manual of Phonology.* Baltimore, MD: Waverly Press.

[forthcoming]Huddleston, Rodney and Geoffrey K. Pullum (forthcoming): *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

[1984]Hughes, G. E. and M. J. Cresswell (1984): *A Companion to Modal Logic.* London: Methuen.

[1998]Immerman, Neil (1998): *Descriptive Complexity.* Berlin: Springer, 1998.

[1977]Jackendoff, Ray S. (1977): $\bar{X}$ *Syntax.* Cambridge, MA: MIT Press.

[1980]Johnson, David E., and Paul M. Postal (1980): *Arc Pair Grammar.* Princeton: Princeton University Press.

[1987]Joshi, Aravind (1987): An introduction to tree adjoining grammars. In Alexis Manaster-Ramer (ed.), *Mathematics of Language*, 87–114. Amsterdam: John Benjamins.

[1996]Keenan, Edward L. and Edward Stabler (1996): Abstract syntax. In *Configurations: Essays on Structure and Interpretation*, ed. by Anna-Maria Di Sciullo (ed.), 329–344. Somerville, MA: Cascadilla Press.

[1998]Keller, Frank (1998): Gradient grammaticality as an effect of selective constraint re-ranking. In M. Catherine Gruber, Derrick Higgins, Kenneth S. Olson, and Tamra Wysocki, eds., *Papers from the 34th Meeting of the Chicago Linguistic Society; Vol. 2: The Panels*, 95–109. Chicago: Chicago Linguistic Society.

[2000]Keller, Frank (2000): *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality.* PhD Thesis, University of Edinburgh.

[2001]Kempson, Ruth; Wilfried Meyer-Viol; and Dov Gabbay (2001): *Dynamic Syntax: The Flow of Language Understanding.* Oxford: Basil Blackwell.

[1999]King, Paul John (1999): Towards truth in head-driven phrase structure grammar. Bericht Nr. 132, Arbeitspapiere des Sonderforschungsbereich 340, Seminar für Sprachwissenschaft, Eberhard-Karls-Universität, Tübingen.

[1990]Kornai, Andràs and Geoffrey K. Pullum (1990): The X-bar theory of phrase structure. *Language* **66**: 24–50.

[1989]Kracht, Marcus (1989): On the logic of category definitions. *Computational Linguistics* **15**: 111–113.

[1993]Kracht, Marcus (1993): Syntactic codes and grammar refinement. *Journal of Logic, Language and Information* **4**: 41–60.

[2001]Kracht, Marcus (2001): Logic and syntax: a personal view. In Michael Zhakharyaschev, Krister Segerberg, Maarten de Rijke and Heinrich Wansing (eds.), *Advances in Modal Logic 2.* Stanford: CSLI Publications.

[1971]Lakoff, George (1971): On generative semantics. In Danny D. Steinberg and Leon A. Jakobovitz (eds.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, 232–296. Cambridge: Cambridge University Press, 1971.

[1984]Langendoen, Terry and Paul M. Postal (1984): *The Vastness of Natural Languages.* Oxford: Basil Blackwell.

[2000]Lasnik, Howard (2000): Syntactic Structures *Revisited: Contemporary Lectures on Classic Transformational Theory.* Cambridge, MA: MIT Press.

[1997]Miller, Philip, Geoffrey K. Pullum, and Arnold M. Zwicky (1997): The Principle of Phonology-Free Syntax: Four apparent counterexamples from French. *Journal of Linguistics* **33**: 67–90.

[1971]Perlmutter, David M. (1971): *Deep and Surface Structure Constraints in Syntax.* New York: Holt Rinehart and Winston.

[1994]Pollard, Carl and Ivan A. Sag (1994): *Head-driven Phrase Structure Grammar.* Stanford, CA: CSLI Publications.

[1999]Pollard, Carl (1999): Strong generative capacity in HPSG. In Gert Webelhuth, Jean-Pierre Koenig, and Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, 281–297. Stanford, CA: CSLI Publications.

[1943]Post, Emil (1943): Formal reductions of the general combinatory decision problem. *American Journal of Mathematics* **65** (1943) 197–215.

[1944]Post, Emil (1944): Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society* **50** (1944) 284–316.

[1993]Prince, Alan and Paul Smolensky (1993): *Optimality Theory* (Technical Report RuCCS TR–2, Rutgers University Center for Cognitive Science). Piscataway, NJ: Rutgers University.

[1997]Pullum, Geoffrey K. and Barbara C. Scholz (1997): Theoretical linguistics and the ontology of linguistic structure. In *1997 Yearbook of the Linguistic Association of Finland*, 25–47. Turku: Linguistic Association of Finland.

[forthcoming]Pullum, Geoffrey K. and Barbara C. Scholz (forthcoming): Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, in press.

[1996]Rogers, James (1996): A model-theoretic framework for theories of syntax. In *34th Annual Meeting of the Assocation for Computational Linguistics: Proceedings of the Conference*, 10–16. San Francisco, CA: Morgan Kaufmann.

[1997]Rogers, James (1997): “Grammarless” phrase structure grammar. *Linguistics and Philosophy* **20**: 721–746.

[1998]Rogers, James (1998): *A Descriptive Approach to Language-Theoretic Complexity.* Stanford, CA: CSLI Publications.

[1999]Rogers, James (1999): The descriptive complexity of generalized local sets. In Hans-Peter Kolb and Uwe Mönnich (eds.), *The Mathematics of Syntactic Structure: Trees and their Logics*, (Studies in Generative Grammar, 44), 21–40. Berlin: Mouton de Gruyter.

[1985]Sag, Ivan A.; Gerald Gazdar; Thomas Wasow; and Stephen Weisler (1985): Coordination and how to distinguish categories. *Natural Language and Linguistic Theory* **3**: 117–171.

[1999]Sag, Ivan A. and Thomas Wasow (1999): *Syntactic Theory: A Formal Introduction.* Stanford, CA: CSLI Publications.

[1996]Schütze, Carson (1996): *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology.* Chicago: University of Chicago Press.

[1986]Thomas, Simon (1986): Theories with finitely many models. *Journal of Symbolic Logic* **51**: 374-376.

[1990]Thomas, Wolfgang (1990): Automata on infinite objects. In J. van Leeuwen (ed.), *Handbook of Theoretical Computer Science*, 135–191. New York: Elsevier Science.