A model is a system that has all the properties of some object of study that are relevant to answering some specific question about it. For example, the mouse is a model of the human being for certain specific purposes of medical inquiry. In philosophical logic, a model is a mathematical structure that represents all and only the objects and relations that hold in some domain that some formal language (such as first-order logic) can describe, and a model theory is a set of rules that map all possible models onto the values of terms in that language (such as truth).

When we talk about modeling human language, there are at least two distinct aspects to language that we might be talking about. One is what Marr (1977) called “the theory of the computation” and Chomsky (1965) called “competence”—that is, the specification of the set of sentences that constitute any human language, paired with their meanings. Such models must by definition aspire to be sound and complete, generating (or, equivalently, accepting) all and only the sentences of the language. Developing such models generally requires paying attention to low-frequency constructions, out in the “long tail” of the Zipfian power-law distribution that characterizes nearly all language events, from word-frequency to construction-frequency.

The other kind of model is what Marr called the “algorithm” and Chomsky called “performance”—that is, the mechanisms whereby a human or an automaton can obtain one of those meanings from the corresponding sentence in a given language, or vice versa. Such models must by definition be approximate, and are often probabilistic or information-theoretic (Shannon and Weaver 1949), with a consequent emphasis on what I assume should be called the “short head” of frequent events in the Zipfian distribution.

The questions associated with these two kinds of models are quite different. If we are talking about competence, then reasonable questions are “is context-free grammar a model for all human grammars?”, and “what is the relation between linguistic categories and cognitive concepts?” If we are talking about performance, then reasonable questions are “does the human parsing mechanism search bottom-up, starting with words?”, and “is that search guided by anything like a probabilistic parsing model?”.

Strictly speaking, these two kinds of questions are independent. It could in theory be the case that the performance mechanism is entirely independent of the specification of abstract competence. (For example, the former could be a finite-state Markov process that merely approximates a more expressive competence grammar.)

However, the fact that competence grammar is so closely tied to to the construction or realization of meaning, and the self-evident fact that competence and performance must have evolved together, as a package-deal, means that any approximation must be close enough to competence grammar to support composition of meaning. (While there have been proposals for “latent” semantics, based on “trajectories” through vector spaces of collocation-based word-meanings, it is far from clear how such representations can do the most basic jobs of linguistic meaning, such as referring to the world or a model.)

Thus, the most promising mechanisms are those which use rules of competence grammar quite directly in a semantically “surface-compositional” derivation. Such mecha-
nisms have the helpful property that they allow us to explain how children can easily learn language from sentences of their language paired with (possibly noisy, possibly ambiguous) language-independent meanings, in a process that resembles the automatic induction of parsers from treebanks (Collins 2003), except that the trees are unordered language-independent meaning representations, and the parsing model is a model over the entire space of possibilities permitted by universal competence. Such surface compositional grammars also allow us to assume that the language-independent semantics is rather closely related to prelinguistic cognition—specifically, sensory-motor cognition of the kind that Piaget (1936) anatomized as culminating in tool-use, approximately coinciding with the onset of language in the child.

The history of the study of language and cognition since these principles were established can be summarized briefly as follows:

From the 1960s until the mid 1970s, there was almost complete unanimity among theoretical linguists, psycholinguists, and computational linguists, concerning the models for the object of study. This consensus was founded on some formal results due to Chomsky, showing that competence could not be exactly captured using finite-state machines or even context-free grammars. The agreed competence model was transformational generative grammar, which the linguists developed, the computational linguists found elegant algorithms to parse, and the psychologists used as a basis for the empirical study of human processing. The agreed performance model was based on the pursuit of a single syntactic analysis, under the guidance of parsing “strategies” (amounting to ordering on rules) to deal with the ambiguity in the competence grammar, supplemented by backtracking or “reanalysis” in cases (by assumption, rare) where the strategies led the processor into a blind alley. There was considerable interest in the long tail of rare events like garden-path sentences, parasitic gaps and inverse quantifier scope.

This consensus was immensely productive, leading to important insights into the nature of the processor, the interactions among modules including syntax, semantics, and context, and to a number of ingenious behavioral and physiological measures of transient processing load that are still in use today.

The consensus soon entirely fell apart, largely because of early disagreements about the role of semantics in the competence theory, and the realization of the unconstrained power of structure-dependent transformational rules. Many contemporary theoretical syntactic accounts offer very little that psychologists and computational linguists can use. Other more semantically- or psycholinguistically-oriented theories of language either allow arbitrarily powerful transformation-like rules in syntax, or talk in terms of global properties and constraints whose relation to specific formal or computational models is left unspecified. As a result, many psychologists have become almost entirely agnostic about the competence theory, talking either in terms of global constraints and strategies unrelated to any specific theory of grammar, or in some extreme cases entirely denying the psychological relevance of linguistics and the competence-performance distinction itself.

Meanwhile, computational linguists have mostly reverted to finite-state and context-free approximations to human language, using parallel-searching bottom-up algorithms and machine-learnable probabilistic parsing models to deal with the huge grammars and
proliferating analyses that are needed for robust practical applications on a large scale, and generally ignoring the long tail.

I don’t think a (psycho-) linguistics in this fragmented state has much in the way of models to offer to the study of cognition. (We ask for bread. They give us empty categories.)

On the other hand, I do not believe that cognitive science has come up with anything whatsoever in the way of an alternative model that addresses the real questions that were raised in that glorious dawn of the formal study of language. Since the time of Lashley (1951) and Miller, Galanter and Pribram (1960), it has been standard to assume that language is closely related in both evolutionary and developmental terms with embodied sensory-motor cognition (in which I include a certain amount of social cognition). There is much talk these days of the “emergence” of language on this basis, often with appeal to evidence from mirror-neurons (??), but the mere observation that chemistry is emergent from physics does not absolve us of the responsibility of formally explaining exactly how the laws we observe in one domain follow from those in the other (Holland 1998). The fact is that we have absolutely no idea how sensory-motor cognition is represented, and it is quite unlikely that the way in which our robots embody it is at all similar to the way we ourselves do.

This paper argues instead for a return to a unified theory of grammar that captures robust linguistic generalizations about universal grammar, while also supporting robust computational language processing and a realistic model of human language processing. The theory in question is strongly lexicalized, in the sense that the lexicon is the only locus for language-specific information. Lexical categories specify the entire local syntactic domain and the associated semantic interpretation for all bounded constructions, including language-specific linear order. A universal syntactic projection mechanism using combinatory rules of low (near-context-free) expressive power governs the assembly of lexical elements into constituents and sentences, establishing all semantic dependencies (including unbounded dependencies) directly. This mechanism is type-dependent, rather than structure-dependent, and is entirely monotonic, never revising structural representations. The semantics is essentially proof-theoretic, oriented towards establishing entailment, rather than model-theoretic in the sense of (and much subsequent work (of course, a model theory must eventually be provided as well, to prove correctness).

The paper will take this theory almost entirely as read (Steedman 2000, passim), except to show that the syntactic and projection mechanism is principally defined in terms of two kinds of type-driven operators corresponding semantically to composition (of functions) and type-raising (of arguments). These operators correspond to two of the simplest combinators that Curry and Feys (1958) used to define applicative systems such as the $\lambda$-calculus. (There is a third combinator that I will ignore here.) Building on Steedman 2002, the paper will argue that such a system could come into being at minimal evolutionary cost, given the prior establishment of a cognitive system, evolved over a much longer period, able to deal with planning of the complexity associated with the sixth and final Piagetian phase of sensory-motor cognitive development.

In particular, the paper relates composition to Lashleyan seriation of actions in plans, and type-raising to Gibsonian affordance of those actions by objects or situations. The
emergence of language in such a system depends on a number of further cognitive developments, including natural-kind concepts for tools (reflected in (some) languages by genders or classifiers), and the concept of agency (reflected in (all) languages as case). (There is evidence to suggest that other primates than humans have access to this much.) The emergence of language also appears to depend on access to certain types of knowledge as affordances of other minds, which the evidence suggests primates other than humans may not have. The form of the latter concepts suggests a semantic origin for recursion as a distinctive property of human language (Hauser, Chomsky and Fitch 2002).

References