

Constructions on holiday

JEFFREY LIDZ and ALEXANDER WILLIAMS*

Abstract

Constructions at Work provides many new empirical observations and sound criticisms of other work. These fail, however, to support Goldberg's two main arguments. They do not undermine the need for innate linguistic structure; nor do they decide for Construction Grammar in particular.

Keywords: learning; Construction Grammar; Generative Linguistics; nativism.

1. Introduction

The set of utterances to which any child acquiring a language is exposed is equally compatible with many distinct descriptions. And yet children converge to a remarkable degree on a common grammar, with agreement on indefinitely many sentences that are novel. Mainly for this reason, Chomsky proposed that the child brings prior biases to the task. Relative to these innate biases, which restrict or structure the learner's range of choices, the input is more decisive, making convergent learning possible. The defining goal of the Chomskyan enterprise has been to identify these biases. In practice it has been agreed that some of these are specific to language acquisition, and some are, in addition, constraints on the format of syntactic representation.

Adele Goldberg is disappointed by this conclusion, and *Constructions at Work (CW)* purports to bring a message of hope:

* University of Maryland Department of Linguistics. Authors' e-mails: <jlidz@umd.edu>; <alxndrw@umd.edu>. Thanks to Tonia Bleam, David Embick, Norbert Hornstein, Bill Idsardi, Michael Israel, Colin Phillips, Anoop Sarkar, and Matt Wagers for helpful discussion during the preparation of this article.

Surely it is premature to give up hope that humans, with our rich cognitive abilities, complex social skills, predilection to imitate, and 100-billion-neuron brains, can learn language from the available input. (Goldberg 2006: 69)

The main aim of the book is, therefore, to show that language acquisition requires few innate biases specific to language, if any. There is no acquisition device specific to language, and, correspondingly, no immanent theory of syntax. The structure of the acquired grammar derives instead from the statistical character of a domain-general learning procedure, resource pressures on rapid processing, principles of rational communication, and native schemes for conceptualizing nonlinguistic experience. Secondly, *CW* implies that Goldberg's brand of Construction Grammar (CxG) is useful in deriving these conclusions.

The arguments in *CW* fail to achieve this aim. The book gathers many important observations, but mostly these do not answer any of the relevant questions. The suggestion that they do is, in our view, based on an oversimplified interpretation both of the observations and of the questions. This weakness is expressed in *CW*'s understanding of learning, of the data that a theory of syntax needs to answer for, and of the kinds of arguments that decide between grammatical architectures.

2. Learning and representation

CW presents two kinds of arguments, (A) and (B), against the assumption that language acquisition requires constraints specific to language, either on the learning device or on its output representations.

- (A) Language acquisition uses some domain-general techniques of statistical learning.
- (B) Some abstract patterns are learned more easily than we might expect.

The implicit promise is that, with continued research, "some" will be replaced by "only" in (A) and "all" in (B). But even the two weaker claims here are compromised by oversimplification in the characterization of what it means to learn and of what is learned.

It is a major theme of *CW* that language acquisition is like categorization tasks generally: they exhibit the same sensitivity to statistical structure in the input. From this Goldberg wants to conclude that there are no innate constraints on linguistic representation. But learning *anything* from probability distributions is impossible without a prior representational space (Shannon 1948, Sober 1994, Tenenbaum and Griffiths 2001, Gallistel 2007). The function of a representational bias is to make the

available statistics informative enough that convergent learning is possible (Fodor 1966, Pinker 1989, Yang 2002, Pearl 2007). So to challenge the Chomskyan program, one would need to show that this bias can itself be induced statistically, using nonlinguistic priors. And *CW* does not do this.

An example of this error is Goldberg's reference to Alishahi and Stevenson 2005, which shows that a Bayesian learning algorithm can link semantic with syntactic representations (cp. Zettlemoyer and Collins 2005). But this demonstration *begins* with these representations.¹ It does not say how a learner might construct them on the basis of distributional generalizations over strings and their contexts of use—exactly where the generativist calls on innate guidance.

As an initial illustration of (B), *CW* discusses an experiment on learning by honeybees (Giurfa 1996). Bees were trained to find food in a Y-shaped maze. At the entrance was a colored stimulus. At the fork in the Y, paired with the two directions, were two more stimuli, only one of which matched the color of the first. The bee was rewarded with food only if she traveled down the fork with the matching stimulus. After this training, the bees entered a new maze of the same type, but where the stimuli matched or contrasted not in color, but in pattern (e.g., horizontal versus vertical stripes). The bees then chose the matching pattern over 75% of the time.

In Goldberg's view, honeybees “learn[ed] the abstract concept of ‘sameness’ and ‘difference’.” (p. 69). Because even insects could learn such an abstract concept, the argument goes, we should not underestimate the human ability to learn abstract concepts from experience, even in language.

But the honeybees did not “learn the abstract concept of sameness.” They learned that sameness of stimuli leads to food. And presumably they did this (as we would) by using a modality-general concept of sameness, ‘same-in-X,’ that was already in place. Had this not been in place, they could only have learned that matching-colors means food, and would therefore have failed in the new maze. Their success shows that their training experience was represented instead in terms of a general sameness relation, with the mode of sameness as a variable parameter. And insofar as this representational preference is explained by no aspect

1. “We assume that the input to the argument structure acquisition process consists of pairs of representations, one for the perceived utterance (what the child hears), and one for the relevant aspect of the observed scene (the semantics described by the utterance).” (Alishahi and Stevenson 2005: 98).

of their experience in the first maze, it expresses the sort of learning bias that poverty-of-the-stimulus arguments aim to uncover. Thus Goldberg's discussion of the bees illustrates the pervasive motivation for the nativist methodology, and the unreliability of her own.

In discussions of human learning, *CW*'s conclusions are similarly weakened by simplifications in the description of what is learned—simplifications we cannot afford when the question is how much it takes to learn exactly what is learned.

For example, Goldberg discusses experiments wherein English-speaking children learn a designed association between the form NP_1 - NP_2 - V , with V a nonce verb, and scenes in which NP_1 's referent appears in the area of NP_2 's, with manner of appearance varying with choice of V . For instance, “the rabbit the hat mooped” was associated with a scene in which a rabbit appears on a hat.

Evidently the children learned a relation between some general description of a string and some general description of a scene. But which descriptions? Goldberg presumes they represent NP_1 and NP_2 as, respectively, the subject and object of a clause whose verb is V , and the referent of NP_2 as bearing a “locative” relation to the V event. Since *SOV* order is not English, and locatives are normally expressed as obliques, and not as direct arguments, Goldberg concludes that children can quickly learn even abnormal associations of form and meaning. But this conclusion is moot without evidence for the presumed representations. For all we know, the children instead regard NP_1 as a fronted object, NP_2 as a subject, and understand the latter as a *source*. In that case the result is like “The hat produced the rabbit,” here shifted to “the rabbit, the hat produced.” We don't even know that they treat the string as a clause, much less one of English. All we know is that they learn to relate two descriptions. And this says nothing about whether the “linking” between grammatical and thematic relations, let alone the format of their respective representations, is innately constrained.

3. English grammar

As a second part of its plan to show how “humans . . . can learn language from the available input,” *CW* endeavors to revise our assumptions about what is learned. This effort is framed in opposition to the “mainstream generativist,” who says that there is some fact Φ , that Φ expresses a principle of syntax, and that this principle is not determined by the input. Goldberg responds that Φ is not a fact, or that it is explanatorily related to something outside syntax. Then the claim is that, subsequent to her redescription, the data available to the learner is richer, with a struc-

ture that makes acquisition possible without the “generativist’s” innate constraints.

Some of Goldberg’s criticisms of specific analyses are sound, and will be widely welcomed. But typically these critiques conflict with no basic “generativist” principle. And too often, Goldberg’s own reanalysis of the data is unsatisfactory, leaving open many questions that the “generativist” rightly considers urgent.

In the last third of *CW*, the polemic opens on three fronts: subject-auxiliary inversion (SAI) in English, extraction islands, and argument structure. Several of the criticisms in the domain of argument structure will not be very controversial; most of the others have been responded to elsewhere (Lidz and Gleitman 2004). Here we address the remaining two topics.

As every English grammarian knows, SAI typically obtains in sentences used for purposes other than a positive assertion:

- (1) a. Had she been singing?
- b. Had she been singing, we would’ve heard her.
- c. At no time had she been singing.

What Goldberg adds is that this generalization “motivates” its form. Since finite “auxiliaries carry information about polarity” (*CW*, 178), fronting of a finite auxiliary is an apt way of “convey[ing] that the polarity involved is not the canonical, positive polarity” (*CW*, 180)—or, that the sentence has irrealis mood. Cases of SAI with “canonical, positive polarity,” like (1d), are deemed “non-prototypical” and therefore free from the proposed “motivation”.

- (1) d. So loudly had she been singing that we couldn’t hear Mick.

Never mind that English auxiliaries do not themselves indicate mood, and that none of the auxiliaries in (1) explicitly mark polarity. We are still left asking why it is the highest auxiliary that is tensed, and consequently fronted. Why not tense and front the lowest auxiliary?

- (2) a. *Was she have singing?
- b. *At no time was she have singing.

It is right to respond that tense always associates with the highest verb, even outside SAI. But to state this as an additional fact—it is the tensed verb that inverts and, coincidentally, it is the first verb that bears tense—is to miss a generalization. Both SAI and association with tense are local processes, targeting the highest head of the right sort. And this locality is exhibited by many constructions, in many languages. Nearest to home, it

is exhibited in German verb fronting constructions (where only the highest verb can tense and front), though these plainly lack the function that Goldberg says is “prototypically” associated with SAI in English.

This locality condition is described in the “Head Movement Constraint” (Travis 1984), or HMC, which entails that one head cannot leapfrog over another. Because SAI obeys this general constraint, “generativists” do not regard it as a *sui generis* construction, one specifically of inverting the subject and the tensed verb. Instead it is viewed as an instance of a general process of head-raising, one that happens to occur (among other places, and for whatever reasons) in sentences used to make speech acts other than positive assertions.

The hypothesis that SAI is head-raising to, in particular, a silent complementizer position directly captures some further facts that are accidental for Goldberg. The tensed verb occupies a position where the language puts its complementizers—rather than, say, being extraposed to a clause final focus position, (3). The overt presence of a complementizer bleeds SAI, (4). And since SAI is movement to a head position, not just to any clause-initial position of prominence, fronted auxiliaries occur lower than fronted phrases, (5).

- (3) *She been singing has?
- (4) *If had I known she would be singing,
- (5) *Has which aria she been singing?

If one still wants “motivation” for the movement, it is worth observing that English complementizers, unlike English auxiliaries but like complementizers in many other languages, do indicate the grammatical mood of their complements.

What is more important, since it has been less clear how children learn it (Chomsky 1975, Crain and Nakayama 1987), is the fact that SAI targets the auxiliary that is structurally highest, (6a), rather than the one that is linearly first, (6b). Why should this be?

- (6) a. Have the people who were sick died?
- b. *Were the people who sick have died?

Perhaps it seems natural to assume that fronting of an expression X can serve to indicate a property only of a clause in which X is an immediate constituent. But this assumption is false. Movement of an embedded wh-phrase to the front of a clause serves to indicate that that clause is interrogative, for example, as is evident when interrogativity is not signaled by SAI:

- (7) Al wonders who we suspect stole the painting.

- (8) Wen glaubst du dass er unter Verdacht hat?
 Who believe you that he under suspicion have?
 ‘Who do you think that he suspects.’ (German)

So we can’t see how (6) follows from Goldberg’s functional motivation for SAI. Yet the “generativist” has available the same answer as before. The HMC dictates that head movement is local. The only auxiliary that can raise to a complementizer position is the one that heads the clausal complement of that complementizer.

Constraints like the HMC, which cut across semantically unrelated constructions, do not grab Goldberg’s attention, since she presumes that shared properties of form stem from shared properties of meaning. Yet such constraints preoccupy “generativists,” since their apparent independence from meaning raises the likelihood that they express unlearned constraints on form. In the case of the HMC, the generativist has three reasons to bet that it is a consequence of universal grammar. It is robustly expressed in a diverse range of languages, in semantically unrelated contexts. It is similar to constraints on other nonlocal dependencies, such as the constraint expressed in (9), where a Raising dependency cannot leapfrog one subject (John) over another (it).

- (9) a. It seems to be likely that John will win.
 b. It seems that John is likely to win.
 c. John seems to be likely to win.
 d. *John seems that it is likely to win.

And it is hard to see how the facts described by the HMC could follow from factors not specific to language. But contrary to what *CW* insinuates, this is a bet that the “generativist” is happy to lose. For if (1–9) follow from something other than innate linguistic constraints, it simplifies the task of describing what those constraints are. No principle of “generativism” requires that every judgment of acceptability have a syntactic explanation.

Thus *CW* fails to engage the observations that motivate the “generativist” position. To say that SAI distorts the canonical word order and signals a noncanonical mood contributes little to our understanding of why the form has just the properties that it does, rather than any others, and how this form relates to other observations about tense, mood, complementizers and nonlocal dependencies in general.

Goldberg’s discussion of Islands has similar weaknesses. It begins helpfully, with a reminder that the acceptability of an extraction dependency often seems to depend on the information-structural properties of the extraction site. In her summary: “Backgrounded constructions,” whose

content is neither topical nor asserted, “are islands” (*CW*, 135). It is right to highlight these facts. And perhaps they have an explanation which convinces the “generativist” that many of the classic Islands are not a pure expression of syntactic constraints.

But no one will be satisfied with the explanation offered by Goldberg:

- (10) “Elements in unbounded dependencies are positioned in discourse prominent slots. *It is pragmatically anomalous to treat an element as at once backgrounded and discourse prominent.*” (*CW*, 135, italics Goldberg).

There are two obvious problems with this. First, relativization is an unbounded dependency that is sensitive to islands, but the head of a relative clause has no intrinsic discourse prominence.

- (11) a. I lost the book that (you thought that) the professor wrote.
b. *I lost the book that I met the professor who wrote.

Second, there is nothing “pragmatically anomalous” about putting focus in a presuppositional (hence “backgrounded”) context, (12), even though focus is a type of discourse prominence.

- (12) I certainly did not read the book that CHOMSKY recommended.

Thus the theoretical consequences of the pragmatic observations remain unclear, since their proposed explanation is faulty.

In addition, there remains a residue of islands (Sprouse 2007) that are not backgrounded contexts, as in extraction out of just a single conjunct.

- (13) *Who did the department give a stipend to Mel and to?

Citing Kluender, Hawkins, and others, Goldberg is hopeful that much of this residue will follow from constraints, not on grammar, but on processing. Maybe. But this program faces serious challenges, including those described in Phillips 2006; see also Phillips and Wagers (2007).

4. Construction Grammar as a grammatical theory

Goldberg might still be right to prefer Construction Grammar over other syntactic theories. Certainly one reads *CW* expecting to see some data that call decisively for the resources of CxG. But this expectation is frustrated.

CxG shares many central premises with other familiar theories, such as LFG, HPSG, and CCG (Steedman 2000). Most importantly, these theories eschew silent lexical items, traces in particular, and are nontransformational. *CW* suggests that these choices make learning or processing

easier by bringing grammatical analysis closer to the surface. But the book includes no direct discussion of how. This is disappointing, but not surprising. There are no decisive demonstrations that any of these assumptions necessarily simplify processing or learning. Some transformational theories are computationally tractable (Joshi, Vijay-Shanker and Weir 1991, Stabler 2003), while some nontransformational theories are not (Berwick 1981, Johnson 1988, Ristad 1990). And sometimes a transformational grammar is more compact, perhaps making it easier to learn (Stabler 1998).

We might expect that *CW* would at least demonstrate the virtues of *constructionism*, a signature feature of CxG. This is the claim that grammatical structure itself contributes substantive predicates to the meaning of a phrase. Constructionism contrasts with the lexicalist claim that all predicates project from lexical signs. Consider (14). It asserts some relation of the table, one which implies that it was smacked. The lexicalist says that this relation is contributed wholly by the verb *smack*. The constructionist allows that it's contributed at least in part by the structural context.

(14) Nick smacked the table.

In defense of constructionism, *CW* observes (Goldberg 1995) that there are associations between a syntactic frame and a semantic relation, not specific to any verb occurring in the frame. It then argues that such associations are readily learned. But none of this decides between constructionism and lexicalism (Pinker 1989, 261ff; Lidz, Gleitman and Gleitman 2004). Both lexicalists and constructionists can recognize such associations. The difference lies just in whether the association is stated over primitives for syntactic analysis, or primitives for lexical analysis (McCawley 1971). Neither the fact that argument structures generalize across verbs, nor the suggestion that they can be learned, provide any evidence for constructionism *per se*.

In particular they provide no evidence for the hybrid constructionism of Goldberg. For her, (14) involves the transitive construction, which assigns to its subject and object general Agent and Patient relations (1995: 117). The *pure* constructionist would stop here, giving (14) the semantics in (15). But for Goldberg the verb itself contributes additional specific thematic relations, as in (16) (*CW*, 38ff).

(15) $\exists e . \text{smack}(e) \ \& \ \text{Agent}(e, \text{Nick}) \ \& \ \text{Patient}(e, \text{table})$

(16) $\exists e . \text{smack}(e) \ \& \ \text{Agent}(e, \text{Nick}) \ \& \ \text{Patient}(e, \text{table}) \ \& \ \text{Smacker}(e, \text{Nick}) \ \& \ \text{Smackee}(e, \text{table})$.

Under many semantic theories (Dowty 1989, Landman 2000), these additions would be redundant: the Agent of a smacking is necessarily its smacker. But a redundant analysis is not necessarily false.

This difference between pure and hybrid constructionism is interesting. For when constructionism is entertained in the “generativist literature,” it is presumed to be pure. Arguments for a constructionist treatment of (particular) thematic relations have all been arguments *against* having these project from the verb (e.g., Carlson 1984, Schein 1993, Kratzer 1996, Marantz 1997, Borer 2003, Pietroski 2005, Williams 2007). So if any of these are right, they invalidate not only the lexicalist analyses that were their intended opponents, but also Goldberg’s hybrid constructionism.²

In sum, while there are ways to distinguish pure constructionism from lexicalism or hybrid constructionism empirically (Schein 1993, Kratzer 2000), Goldberg misses the chance to pursue these, leaving us with no justification for her preferences.

5. Conclusion

How far can language learners get with “the available input” and general intelligence? Goldberg’s answer is based in several hypotheses that we are sympathetic to. Some facts taken to have a syntactic explanation may find a deeper explanation elsewhere; some of the biases that assist the learner may not be innate constraints on syntactic representation; constructionism may be true. But *CW* simply fails to establish any of this. There is not one case where *CW* successfully demonstrates that general intelligence, functional and performance pressures, and domain general statistical learning are sufficient to explain a grammatical pattern, or the acquisition of a linguistic ability. And therefore Goldberg fails to establish her own answer to the motivating question.

Received 9 November 2007

University of Maryland

Revision received 30 January 2008

2. Schein’s (1993, Ch. 4) argument may be the strongest of these. He discusses sentences where subject and object quantifiers appear to be mutually independent, with neither in the scope of the other. But, if the verb itself introduces these arguments, i.e., its interpretation includes variables relating to both, then this independence is arguably impossible. A quantifier that binds one of the two variables will have the other in its scope.

References

- Alishahi, Afra and Suzanne Stevenson
 2005 A probabilistic model of early argument structure acquisition. *Proceedings of the 2nd Annual Conference of the Cognitive Science Society*, 97–102. Stresa, Italy.
- Berwick, Robert C.
 1981 Computational complexity of lexical functional grammar. *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics*, 7–12.
- Borer, Hagit
 2003 Exo-skeletal and endo-skeletal explanation: Syntactic projections and the lexicon. In Moore, J. and M. Polinsky (eds.), *The Nature of Explanation in Linguistic Theory*. Stanford, CA: CSLI, 31–67.
- Carlson, Greg
 1984 Thematic roles and their role in semantic interpretation. *Linguistics*, 22, 259–279.
- Chomsky, Noam
 1975 *Reflections on Language*. New York: Pantheon.
- Crain, Steven and Mineharu Nakayama
 1987 Structure dependence in grammar formation. *Language* 63, 522–543.
- Dowty, David
 1989 On the semantic content of the notion of ‘thematic role.’ In Chierchia, G. (ed.), *Properties, Types, and Meanings: Volume 2*. Dordrecht: Kluwer, 69–130.
- Fodor, Jerry A.
 1966 How to learn to talk: Some simple ways. In Smith, F. and G. Miller (eds.), *The Genesis of Language*. Cambridge, MA: The MIT Press, 105–122.
- Gallistel, Charles R.
 2007 Learning and representation. In Menzel, R. and J. Byrne (eds.), *Learning and Memory: A Comprehensive Reference*. New York: Elsevier.
- Giurfa, Martin
 1996 Symmetry perception in an insect. *Nature* 382, 458–461.
- Goldberg, Adele
 1995 *Constructions*. Chicago: University of Chicago Press.
- Goldberg, Adele
 2006 *Constructions at Work: The Nature of Generalizations in Language*. Oxford: Oxford University Press.
- Johnson, Mark
 1988 *Attribute Value Logic and the Theory of Grammar*. Stanford: CSLI.
- Joshi, Aravind K., K. Vijay-Shanker, and David Weir
 1991 The convergence of mildly context sensitive grammar formalisms. In *Foundational Issues in Natural Language Processing*. Cambridge: The MIT Press, 31–81.
- Kratzer, Angelika
 1996 Severing the external argument from its verb. In Rooryck, J. and L. Zaring (eds.), *Phrase Structure and the Lexicon*. Dordrecht: Kluwer, 109–137.
- Kratzer, Angelika
 2000 The event argument, Ch. 2. Manuscript available from semanticsarchive.net.
- Landman, Fred
 2000 *Events and Plurality*. Dordrecht: Kluwer.

- Lidz, Jeffrey and Lila Gleitman
2004 Yes, we still need Universal Grammar. *Cognition* 94, 85–93.
- Lidz, Jeffrey, Henry Gleitman, and Lila Gleitman
2004 Kidz in the 'hood: Syntactic bootstrapping and the mental Lexicon. In Hall, D. G. and S. R. Waxman (eds.), *Weaving a Lexicon*. MIT Press: Cambridge, 603–636.
- Marantz, Alec
1997 No escape from syntax. *U. Penn Working Papers in Linguistics* 4.2, 201–226.
- McCawley, James D.
1971 Pre-lexical syntax. In O'Brien, R. (ed.), *Report on the 22nd Annual Round Table Meeting on Linguistics and Language Studies*. Washington, DC: Georgetown University Press, 19–33.
- Pearl, Lisa
2007 Necessary bias in language learning. PhD Dissertation, University of Maryland.
- Phillips, Colin
2006 The real-time status of islands. *Language* 82, 795–823.
- Phillips, Colin and Matthew Wagers
2007 Relating structure and time in linguistics and psycholinguistics. In Gaskell, G. (ed.), *Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press, 739–756.
- Pietroski, Paul
2005 *Events and Semantic Architecture*. Oxford: Oxford University Press.
- Pinker, Steven
1989 *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: The MIT Press.
- Ristad, Eric
1990 Computational structure of GPSG models. *Linguistics and Philosophy* 13, 521–587.
- Schein, Barry
1993 *Plurals and Events*. Cambridge: The MIT Press.
- Shannon, Claude
1948 A mathematical theory of communication. *Bell System Technical Journal* 27, 379–423.
- Sober, Elliot
1994 No model, no inference. In Stalker, D. (ed.), *Grue!*. Chicago and LaSalle, Illinois: Open Court, 225–240.
- Sprouse, Jon
2007 A program in experimental syntax. PhD Dissertation, University of Maryland.
- Stabler, Edward P.
1998 Acquiring languages with movement. *Syntax* 1, 72–97.
- Stabler, Edward P.
2003 Comparing three perspectives on head movement. *UCLA Working Papers in Linguistics* 10, 178–198.
- Steedman, Mark
2000 *The Syntactic Process*. Cambridge, MA: The MIT Press.
- Tenenbaum and Griffiths
2001 Generalization, similarity and Bayesian inference. *Behavioral and Brain Sciences* 24, 629–640.

- Travis, Lisa
1984 Parameters and effects of word order variation. MIT PhD dissertation.
- Williams, Alexander
2007 Patients in Igbo and Mandarin. In Dölling, X. et al. (eds.), *Event Structure in Linguistic Form and Interpretation*. Berlin: Mouton de Gruyter.
- Yang, Charles
2002 *Knowledge and Learning in Natural Language*. New York: Oxford University Press.
- Zettlemoyer, Luke and Michael Collins
2005 Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *Proceedings of the Twenty-first Conference on Uncertainty in Artificial Intelligence*, 658–666.