The learning and emergence of mildly context sensitive languages

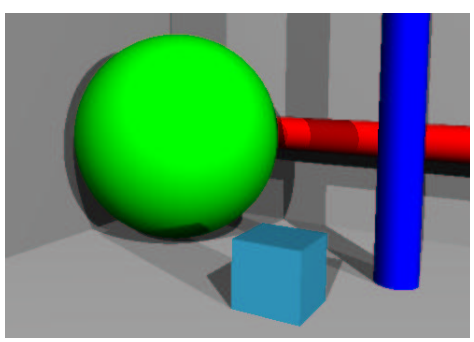
Edward P. Stabler[1], Travis C. Collier[2], Gregory M. Kobele[1], Yoosook Lee[2], Ying Lin[1], Jason Riggle[1], Yuan Yao[2], and Charles E. Taylor[2]

[1] UCLA Linguistics
Box 951543, Los Angeles, CA 90095-1542 USA
[2] UCLA Organismic Biology, Ecology and Evolution
Box 951606, Los Angeles, CA 90095-1606 USA
http://taylor0.biology.ucla.edu/al/

Abstract. This paper describes a framework for studies of the adaptive acquisition and evolution of language, with the following components: language learning begins by associating words with cognitively salient representations ("grounding"); the sentences of each language are determined by properties of lexical items, and so only these need to be transmitted by learning; the learnable languages allow multiple agreements, multiple crossing agreements, and reduplication, as mildly context sensitive and human languages do; infinitely many different languages are learnable; many of the learnable languages include infinitely many sentences; in each language, inferential processes can be defined over succinct representations of the derivations themselves; the languages can be extended by innovative responses to communicative demands. Preliminary analytic results and a robotic implementation are described.

1 Introduction

Human language is learned in context, apparently easily, and not only by the most highly intelligent among us. In fact, human language recognition is 'mandatory' and 'automatic' in a certain sense, like human vision: when you look at a scene like the one below, you cannot help seeing it as having 3D objects in it, objects that can remain constant through significant variations in lighting and orientation (Rock, 1983; Bennett, Hoffman, and Prakash, 1989):



An image like this presents itself on the retina as a large array of color and light intensities, but is interpreted as a small array of objects arranged in 3D space. Similarly, when you hear a sentence like (1), you cannot help hearing it as a linguistic structure of a certain kind (Stroop, 1935; MacLeod, 1991) with familiar objects in it that appear in various forms.

(1) The economy, Bush said, has been improving

There is another important similarity between visual and language perception: just as we cannot help noticing complete objects when parts of them are occluded and discontinuous in the visual image (Kanizsa, 1985; Kanizsa and Gerbino, 1982), (Chomsky, 1956) points out that there are apparently similarly discontinuous presentations of units in human language. In (1), for example, not only is the subject *the economy* separated from the predicate *has been improving* by another subject and predicate, but also *have* and *-en* plausibly form a unit, and so do *be* and *-ing*. We see these same familiar elements in examples like this:

(2) The economy improves

(3) The indexes have fall-en

(4) Expectations are ris-ing

(5) The economy Bush said has be-en improv-ing

Recognizing familiar structures like this – through their discontinuous presentations – is apparently extremely easy for speakers of English, and presents no particular challenge for any normal language learner.

Notice also that while there is noise in any image like the one above (e.g., the surface of the sphere is not perfectly smooth because of rendering and print limitations), even if there were no noise at all (e.g. if the image were given by functions that could deliver arbitrarily precise results about light and color placement), there would still be a challenge in saying how a 3D-like description is computed from the 2D one. There is also noise in language perception (visual or auditory or tactile), but even in a noiseless setting, the challenge of computing the conceptual representation from the percept is non-trivial.

These analogies between vision and language may be more than superficial: in fact, visual and spatial abilities may be important precursors for linguistic abilities (Steedman, 2000), but then there is a puzzle about the plasticity of human language. The ability to recognize visually presented objects is present from the earliest stages and occurs automatically, regardless of the beliefs and goals and "intelligence," general problem solving ability, so much so that early vision is often said to be not cognitive, but only sensory (Pylyshyn, 1999; Kanizsa and Gerbino, 1982). When the recognition of unpredicted events is important to survival, this relative (but of course not absolute) lack of plasticity in vision may improve fitness. Linguistic abilities show a much greater degree of plasticity, but, once acquired, they seem similarly automatic. Are there models of language users that account for this combination of properties?

Recent linguistic and learning results provide a setting in which this and related questions can be studied. The range of patterns and discontinuities found in human languages goes beyond the power of context free formalisms, but not by much: the "mildly context sensitive" (MCS) grammars seem capable of capturing them elegantly, and there are positive learning results that can be extended to them. This paper uses a particular MCS formalism called 'minimalist grammar' (MG) to support a model of language emergence and transmission among agents who already have some (perceptual, maybe even "non-cognitive") abilities for recognizing simple objects from discontinuous presentations. The model has these properties:

i. Dependencies among words can be inferred and grounded in certain settings.
ii. MG languages are specified by their lexicons: i.e. by the association of words with features that specify dependencies.
So the plasticity of language comes from the finitely many learned lexical associations only, not from variations in the basic (possibly "non-cognitive") mechanisms for recognizing composites of these elements.
iii. A certain (infinite) class of MG languages, the "rigid" MG languages, are (provably) learnable from dependency structures.
iv. Recursive language structure emerges quickly and easily in this class.
v. Each MG derivation is unambiguously specified by the sequence of lexical items in it. This provides a natural conceptual representation for reasoning. So on this view, rather than translating linguistic structures into quite different conceptual representations for reasoning, language takes the shape it does partly because that is the shape of objects we reason with.
vi. Language learning can be complete and perfect in principle (provably), but in many actual settings it will be imperfect, and these transmission errors may propagate in a population.

After briefly introducing this model, we describe a robotic realization.

1.1 Grounding

Humans learn language in context. It is natural to assume that learning begins with the identification of correspondences between linguistic tokens and cognitively salient features of the context, where the cognitively salient features will tend to be similar across conspecific participants in a common situation. Many studies have considered this problem (Niyogi, 2002; Vogt, 2000; Kirby, 1999; Steels, 1996). In the present context, we have framed the grounding problem as a learning problem, so that the methods of formal learning theory can be brought to bear. Following (Siskind, 1996), we can regard the evidence available to the learner as a sequence of paired linguistic signals and conceptual representations, where the linguistic signals come from speakers with a certain unambiguous "target" pairing between morphemes and conceptual elements ("sememes"). Even in the ideal setting where all possible elements appear in all possible utterances somewhere in the data – "positive text" in the sense of (Gold, 1967; Jain et al., 1999) – Siskind's algorithm will not always succeed in identifying the target pairing, but (Kobele et al., 2003) provides an algorithm that is guaranteed to succeed in a broad (and defined) range of conditions. Single word and short utterances can be especially valuable in the early stages (Brent and Siskind, 2001), while in later stages the appearance of a new word will often correspond to the appearance of a new sememe, and the association can be confidently established by a single exposure.

Again following (Siskind, 1996), notice that once a unary sememe (a property) has been paired with a symbol P, and a constant sememe (a name) with a symbol a, it is not difficult to guess (and sometimes confirm by observation) that the utterance aP signifies that the thing associated with a by grounding has the property associated with P, and in this case we also obtain the information that the language allows the order Subject-Predicate, and information about how the meaning of this structure is related to the meanings of its parts. In the case where the agent observes two things associated with morphemes a, b in a binary relation associated with some morpheme R, the utterance aRb conveys similar information. In general, with richer assumptions about the syntax of the language, more sophisticated grounding strategies can be explored.

2 Minimalist grammars

The syntax of the present study is inspired by linguistic proposals (Chomsky, 1995; Brody, 1995) that have aggressively stripped redundancy from linguistic structures. (Stabler, 1997) provides a formal model of some of these proposals, and many variants have now been considered (Kobele, 2002; Michaelis, 2002; Frey and Gärtner, 2002; Kobele et al., 2002; Niyogi, 2001; Stabler and Keenan, 2000; Michaelis, Mönnich, and Morawietz, 2000; Lecomte and Retoré, 1999). Most of these systems are weakly equivalent (and strongly similar) to the simplest variant (Stabler and Keenan, 2000), and also to many other independently proposed "mildly context sensitive" grammatical systems (Joshi, 1985; Weir, 1988; Vijay-Shanker and Weir, 1994). These systems define an MCS strict superset of the context free languages (Michaelis, 1998), they are efficiently parsable (Harkema, 2000), and an infinite class of these languages is learnable from semantic structures of a certain kind, as described below.

A 'minimalist grammar' (MG) has two parts: (i) a finite lexicon and (ii) two fixed structure building rules. The **lexicon** is finite set of lexical items, each of which associates a string with an arbitrary finite sequence of features of one of these 4 kinds:

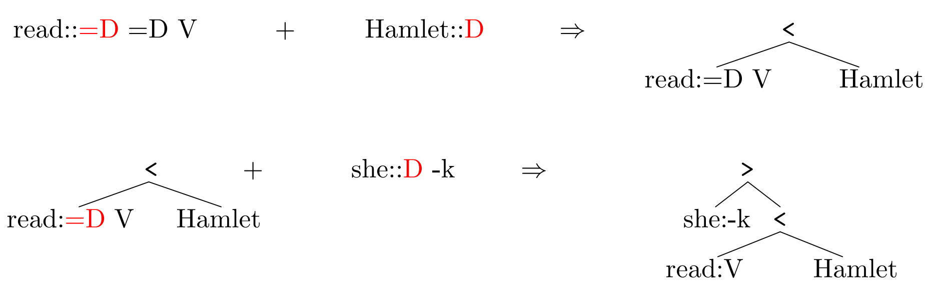
	(examples)
category features	N,V,A,P,C,D,T,...
selector features	=N,=V,=A,=P,=C,=D,=T,...
licensor features	+wh,+k,...
licensee features	-wh,-k,...

So for example, a grammar could have a set of 4 lexical items like this:

G0={ read::=D =D V, Hamlet::D, she::D -k, will::=V +k T }

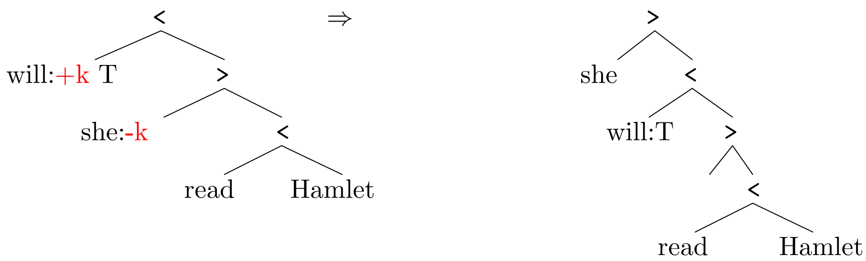
The features of the first item in G0 indicate that it is pronounced "read," and that it must select two expressions with category D ("determiner") to form a complex (a "verb phrase") with category V. The second item "Hamlet" has category D. The third, "she," has category D and -k indicates that it must move to a "case" position. And the fourth item, "will," selects a verb phrase and triggers a case movement to form a T ("tensed") phrase. The notions of selection and movement are defined by the structure building operations.

There are two structure building operations which build trees, and these operations are constant across all grammars: *merge* and *move*. Merge is a binary partial function whose first argument has =F (for some F) as its first syntactic feature, indicating that it wants to select a constituent whose category is F, an operation defined as follows. If the first argument, the selector, is a simple lexical item, then the selected element forms a tree by attaching to the right. If the selector is already a tree, then the second argument attaches to the left:



When merge applies it deletes the pair of features that trigger its action (here, =D and D), building binary tree structures like those shown, with an order symbol at each internal node that "points" towards the "head" of the complex. The features of a tree or subtree are always just the features of its head. And in any tree, any subtree is "maximal" iff no larger subtree has the same head. Notice also that the operations delete features in order, from left to right. The only features affected by merge are those on the heads of the two arguments.

The second structure building operation, move, is a partial unary function on trees. It applies to a tree whose head has a +X feature, if there is exactly one maximal subtree whose head has -X. In that case, that maximal tree with the -X head is moved up to attach at the left edge of the complex as shown. Like merge, this operation deletes a pair of features (here, +k and -k) and creates a discontinuous dependency:



If T is the "start category" of the grammar, then the tree on the right is a completed derivation of the sentence *she will read Hamlet*, since T is the only remaining syntactic feature. Our syntactic representations are sparser than usual, but it is a simple matter to translate derivations in this formalism into the representations that are common in linguistic theory (Stabler, 1997; Stabler, 1999).

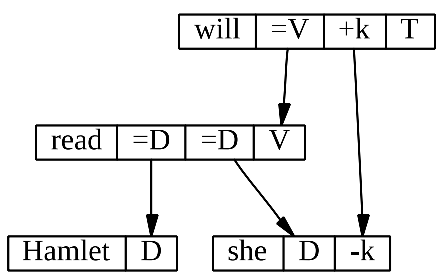
Since merge and move do not vary, a minimalist grammar (MG) is completely specified by its lexicon, like G0 above. A derivation from any such grammar simply involves the application of operations that check features until the only remaining feature is the "start category" T. Numbering the lexical items of G0

G0={ 1.read::=D =D V, 2.Hamlet::D, 3.she::D -k, 4.will::=V +k T }

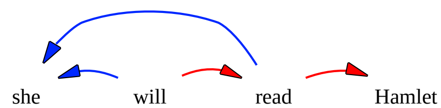
it is easy to see that

$$\text{move}(\text{merge}(4, \text{merge}(\text{merge}(1, 2), 3)))$$

is the tree for *she will read Hamlet* shown just above. That derivation can also be completely specified by saying which feature checking relations obtain among the lexical items, as depicted by the following graph (Stabler, 1999):



Notice that in this graph, each arc is a feature checking relation established in the derivation of the string, but not only that, each arc corresponds to a prominent semantic relation of the sort mentioned in the previous section, except for the "movement" arc that connects +k to -k, and this latter arc is directly evidenced in a change from canonical string position, moving one of the verb's arguments away from the verb. This sets the stage for a dependency-based approach to learning, where the syntactic dependencies either mirror semantic ones or are evidenced by string position. Clearly the learner does not have prior information about the identity of the syntactic features involved in the utterance, but, stripping away these features, we have a dependency structure, a "d-structure", which the learner can plausibly identify (given reasonable assumptions about the nature of the observed utterances):



3 Learning

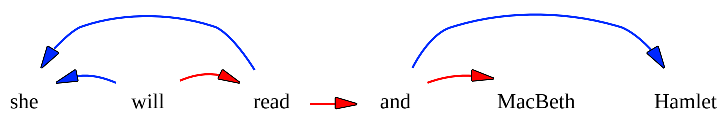
In the simple 'Gold paradigm' model of learning mentioned above, generous assumptions about the data and available computational resources assure that finite languages can be learned, but when the targets come from an infinite class of infinite languages, most of the results are negative. For example, no strict superset of the finite languages is learnable. However, we do not want to learn all of the finite languages, and the languages we want to learn have some structure in common. There is a tradition of positive results of this kind. (Angluin, 1982) shows that an infinite subset of the finite state languages can be learned from example sentences, in the Gold sense of perfect identification. And (Kanazawa, 1998) shows that from some basic dependency structures among the words of the language of a reduced classical categorial grammar, it is possible to identify the target grammar itself (up to alphabetic variants), if the language is *rigid* in the sense that no word has more than one syntactic category.

Kanazawa's results can be extended to MG languages that are rigid in the sense that no two different grounded morphemes (i.e. morpheme-sememe pairs) have the same syntactic roles (Retoré and Bonato, 2001; Stabler, 2002). This rigidity assumption might seem to be a very un-human-like requirement, since many words are ambiguous: *bank* is a noun and a verb. But the rigidity requirement is much less of an imposition when the lexicon is grounded. Then there are multiple elements *bank*: one grounded in financial institutions that is a noun; one grounded in river edges that is a noun; one grounded in a certain financial activity that is a verb. And as we have seen, the learner can begin grounding the language before calculating the syntactic dependencies of its elements. This is the assumption adopted here, and this assumption allows the learner to gather information about a word from multiple utterances, with a straightforward extension of the well-understood process of grammar "unification" from (Buszkowski and Penn, 1990) used by Kanazawa. We describe how the learner proceeds with a simple example.

The learner begins by identifying semantic relations among morphemes (described above) and shifts in the positions of semantically related elements, like the position of subject of in *she will read Hamlet*, (as indicated in the d-structure above). The d-structure is a labeled acyclic graph with a root. Checking the d-structure above, it is easy to see that the root is *will*, so we assign this root the start category T. Then we have various different arcs to other elements which we assume to be ranked in salience, or priority: internal argument<external argument<oblique arguments. Assigning arbitrary categories $A, B, C, \ldots$ to each semantic relation, labeling the end of each arc with a new category and the origin with the corresponding selection feature, and then labeling the end of each string-shifting arc with a new licensee feature and the origin with the corresponding licensor, we obtain a grammar that generates exactly the string given, with the d-structure given, and no other. In this case, the grammar is (up to alphabetic variance):

G1={ read::=A =B C, Hamlet::A, she::B -D, will::=C +D T }

Suppose that the learner now hears the sentence *she will read Hamlet and MacBeth* with the d-structure:



The learner computes a grammar that generates exactly this string, using all new features:

G2={ read::=E =F G, and::=H =I E, MacBeth::I Hamlet::H
she::F -J, will::=G +J T }

Now, if the common words of the two utterances are grounded to the same semantic values, the learner unifies the grammars to obtain a single rigid grammar that can generate both sentences:

G3={ read::=A =B C, and::=I =A A, MacBeth::I Hamlet::A
she::B -D, will::=C +D T }

At this point, the grammar is already recursive in the lexical entry for *and*, and generates an infinite language. This happens because the learner noticed first that the object of *read* has some category A, and then noticed that the object of *read* can be a complex that properly includes another constituent of that same category A. So in this framework, recursion comes from the assumption that two expressions with the same sound and same meaning will play the same role in each utterance, and this can happen with just a few, simple, grounded utterances. (This is a kind of "bottleneck," but it is a very narrow, semantically-based one – cf. Kirby 1999.)

4 Conceptual representation and language generation

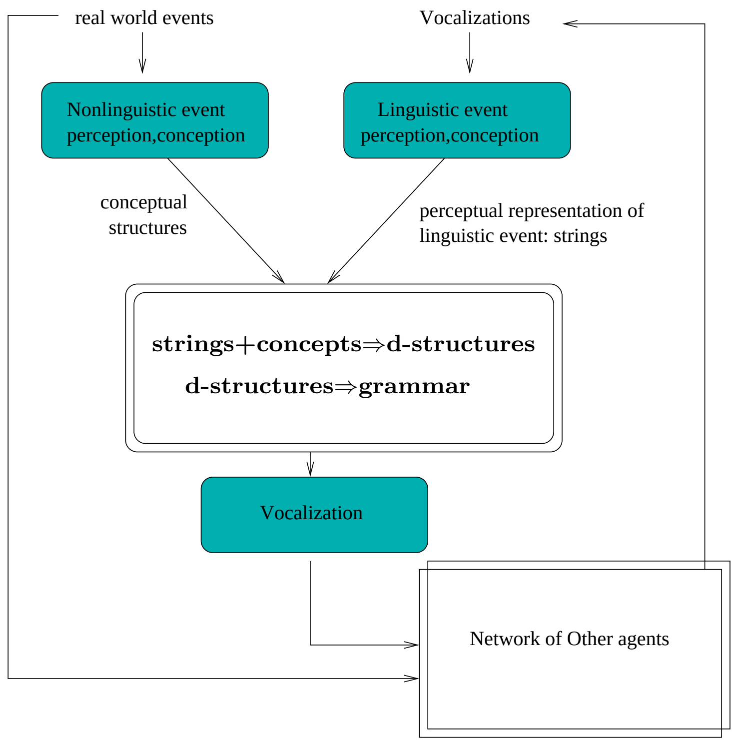
We observed above that $move(merge(4, merge(merge(1, 2), 3)))$ is the derivation of *she will read Hamlet* from grammar G0. It is not hard to show that the sequence of lexical items in this derivation, 4123, has at most one derivation, and this holds regardless of the grammar (Hale and Stabler, 2001). (A given sequence of words, on the other hand, like *she will read Hamlet* can be multiply, even infinitely ambiguous, as in context free languages.) In fact, the sequence 4123 is a Polish notation logic in which the functors of specific arity always precede their arguments. In standard propositional logic, the ambiguous $\neg p \vee q$ corresponds to two unambiguous Polish expressions: $\vee\neg pq$ and $\neg \vee pq$ (Shoenfield, 1967). In the same way, the sequence of lexical items used in a MG derivation, in order, is always unambiguous, compact notation suitable for conceptual representation and inference. Human languages seem to be designed well for monotonicity-based inference methods (Fyodorov, Winter, and Francez, 2003; Bernardi, 2002; Sanchez-Valencia, 1991; Purdy, 1991), rather than the long sequences of modus ponens steps which can achieve the same results, but a careful discussion is beyond the scope of this paper.

The function from lexical sequences (like 4123) to pronouncable strings (like *she will read Hamlet*) can be computed in linear time, and this provides the basis for a theory of language generation ("vocalization"): when a conceptual representation is formed of elements that are grounded to linguistic expressions, the step to the well-formed string is linear; and when some element of a conceptual representation is not associated with a linguistic element, the speaker can choose a semantically related element, or innovate a new expression for the audience to learn. We are currently extending this simple approach to allow generation to be influenced by the speaker's model of the audience.

Previous work on language generation has tackled much more difficult (and sometimes impossible - cf. Shieber 1993) problems, based in part on the idea that the conceptual to linguistic map must be very highly many-to-one. We believe that this component of the problem is over-estimated: by and large, different linguistic expressions correspond to different conceptual representations. Even with this assumption though, many open problems remain (Stone et al., 2001).

5 Robotic realization and future work

The previous sections have briefly introduced the assembly of results that was given in (i-vi) on page 3. Human perceptual and conceptual abilities are quite impressive and still largely unknown, so it is appealing to study simpler artificial systems in an environment that is well understood. We are studying several robotic environments with roughly the configuration diagrammed below. In one study, the "external events" being monitored were network communications, which can be described in a simple English-like language. An early version of this system was reported in (Wee et al., 2001), and the basic linguistic components of the current system are publically available from the project web page (given with the addresses on the first page). In a networked environment now under study, different machines describe their network communications to each other, and since many of these will be common, the machines can, in certain settings, successfully ground their utterances. In this setting, we are exploring several kinds of language innovation: shortening common expressions ('hypocoristics'),



and innovation of new morphemes. The learning model extends immediately to guarantee that other members of the language community will be able to learn these new elements, but the dynamics of these interactions in real settings have not yet been studied. We are also studying extensions of the system to more challenging perceptual environments with more severe grounding problems.

References

Angluin, Dana. 1982. Inference of reversible languages. *Journal of the Association for Computing Machinery*, 29:741–765.

Bennett, B., D. Hoffman, and K. Prakash. 1989. *Observer Mechanics: A formal theory of perception*. Academic, NY.

Bernardi, Raffaela. 2002. *Reasoning with Polarity in Categorial Type Logic*. Ph.D. thesis, University of Utrecht, Utrecht.

Brent, Michael R. and Jeffrey Mark Siskind. 2001. The role of exposure to isolated words in early vocabulary development. *Cognition*, 81:B33–B34.

Brody, Michael. 1995. *Lexico-Logical Form: A Radically Minimalist Theory*. MIT Press, Cambridge, Massachusetts.

Buszkowski, Wojciech and Gerald Penn. 1990. Categorial grammars determined from linguistic data by unification. *Studia Logica*, 49:431–454.

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory*, IT-2:113–124.

Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press, Cambridge, Massachusetts.

Frey, Werner and Hans-Martin Gärtner. 2002. On the treatment of scrambling and adjunction in minimalist grammars. In *Proceedings, Formal Grammar'02*, Trento.

Fyodorov, Yaroslav, Yoad Winter, and Nissim Francez. 2003. Order-based inference in natural logic. *Research on Language and Computation*. Forthcoming.

Gold, E. Mark. 1967. Language identification in the limit. *Information and Control*, 10:447–474.

Hale, John and Edward P. Stabler. 2001. Representing derivations: unique readability and transparency. Johns Hopkins and UCLA. Publication forthcoming.

Harkema, Henk. 2000. A recognizer for minimalist grammars. In *Sixth International Workshop on Parsing Technologies, IWPT'2000*.

Jain, Sanjay, Daniel Osherson, James S. Royer, and Arun Sharma. 1999. *Systems that Learn: An Introduction to Learning Theory (second edition)*. MIT Press, Cambridge, Massachusetts.

Joshi, Aravind. 1985. How much context-sensitivity is necessary for characterizing structural descriptions. In D. Dowty, L. Karttunen, and A. Zwicky, editors, *Natural Language Processing: Theoretical, Computational and Psychological Perspectives*. Cambridge University Press, NY, pages 206–250.

Kanazawa, Makoto. 1998. *Learnable Classes of Categorial Grammmars*. CSLI Publications/FOLLI, Stanford, California.

Kanizsa, G. 1985. Seeing and thinking. *Acta Pschologica*, 59:23–33.

Kanizsa, G. and W. Gerbino. 1982. Amodal completion: seeing or thinking? In B. Beck, editor, *Organization and Representation in Perception*. Erlbaum, Hillsdale, NJ, pages 167–190.

Kirby, Simon. 1999. Learning, bottlenecks, and the evolution of recursive syntax. In E.J. Briscoe, editor, *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*, Cambridge. Cambridge University Press.

Kobele, Gregory M. 2002. Formalizing mirror theory. *Grammars*, 5:177–221.

Kobele, Gregory M., Travis Collier, Charles Taylor, and Edward Stabler. 2002. Learning mirror theory. In *6th International Workshop on Tree Adjoining Grammars and Related Frameworks*.

Kobele, Gregory M., Jason Riggle, Travis Collier, Yoosook Lee, Ying Lin, Yuan Yao, Charles Taylor, and Edward Stabler. 2003. Grounding as learning. In *Language Evolution and Computation Workshop, ESSLLI'03*.

Lecomte, Alain and Christian Retoré. 1999. Towards a minimal logic for minimalist grammars. In *Proceedings, Formal Grammar'99*, Utrecht.

MacLeod, Colin M. 1991. Half a century of research on the Stroop effect: An integrative approach. *Psychological Bulletin*, 109:163–203.

Michaelis, Jens. 1998. Derivational minimalism is mildly context-sensitive. In *Proceedings, Logical Aspects of Computational Linguistics, LACL'98*, NY. Springer.

Michaelis, Jens. 2002. Note on the complexity of complex heads in a minimalist grammar. In *Proceedings of the 6th International Workshop on Tree Adjoining Grammars and Related Frameworks, TAG+6*.

Michaelis, Jens, Uwe Mönnich, and Frank Morawietz. 2000. Algebraic description of derivational minimalism. In *International Conference on Algebraic Methods in Language Proceesing, AMiLP'2000/TWLT16*, University of Iowa.

Niyogi, Sourabh. 2001. A minimalist implementation of verb subcategorization. In *Seventh International Workshop on Parsing Technologies, IWPT'2001*.

Niyogi, Sourabh. 2002. Bayesian learning at the syntax-semantics interface. In *Procs., 24th Annual Meeting of the Cognitive Science Society*, pages 697–702.

Purdy, William C. 1991. A logic for natural language. *Notre Dame Journal of Formal Logic*, 32:409–425.

Pylyshyn, Zenon. 1999. Is vision continuous with cognition? the case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22(3).

Retoré, Christian and Roberto Bonato. 2001. Learning rigid Lambek grammars and minimalist grammars from structured sentences. In L. Popelínský and M. Nepil, editors, *Proceedings of the Third Learning Language in Logic Workshop, LLL3*, pages 23–34, Brno, Czech Republic. Faculty of Informatics, Masaryk University. Technical report FIMU-RS-2001-08.

Rock, Irvin. 1983. *The Logic of Perception.* MIT Press, Cambridge, Massachusetts.

Sanchez-Valencia, V. 1991. *Studies on Natural Logic and Categorial Grammar.* Ph.D. thesis, University of Amsterdam, Amsterdam.

Shieber, Stuart. 1993. The problem of logical form equivalence. *Computational Linguistics*, 19:179–190.

Shoenfield, Joseph R. 1967. *Mathematical Logic.* Addison-Wesley, Menlo Park, California.

Siskind, Jeffrey Mark. 1996. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61:39–91.

Stabler, Edward P. 1997. Derivational minimalism. In Christian Retoré, editor, *Logical Aspects of Computational Linguistics.* Springer-Verlag (Lecture Notes in Computer Science 1328), NY, pages 68–95.

Stabler, Edward P. 1999. Remnant movement and complexity. In Gosse Bouma, Erhard Hinrichs, Geert-Jan Kruijff, and Dick Oehrle, editors, *Constraints and Resources in Natural Language Syntax and Semantics.* CSLI, Stanford, California, pages 299–326.

Stabler, Edward P. 2002. Structures for learning. In *CoLogNet Lecture, ESSLLI'02*, Trento. Publication forthcoming.

Stabler, Edward P. and Edward L. Keenan. 2000. Structural similarity. In A. Nijholt, G. Scollo, T. Rus, and D. Heylen, editors, *Algebraic Methods in Language Processing, AMiLP 2000*, University of Iowa. Revised version forthcoming in *Theoretical Computer Science.*

Steedman, Mark J. 2000. *The Syntactic Process.* MIT Press, Cambridge, Massachusetts.

Steels, Luc. 1996. Synthesizing the origins of language and meaning using co-evolution, self-organisation and level formation. In J. Hurford, C. Knight, and M. Studdert-Kennedy, editors, *Evolution of Human Language*. Edinburgh University Press, Edinburgh, pages 161–165.

Stone, Matthew, Christine Doran, Bonnie Webber, Tonia Bleam, and Martha Palmer. 2001. Microplanning with communicative intentions: The spud system. Technical Report CL/0104022, arXiv.org e-Print archive, `http://arXiv.org/`.

Stroop, J. Ridley. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 12:643–662.

Vijay-Shanker, K. and David Weir. 1994. The equivalence of four extensions of context free grammar formalisms. *Mathematical Systems Theory*, 27:511–545.

Vogt, Paul. 2000. *Lexicon grounding on mobile robots*. Ph.D. thesis, Vrije Universiteit Brussel, Laboratorium voor Artificiële Intelligentie.

Wee, K., T. Collier, G. Kobele, E.P. Stabler, and C. Taylor. 2001. Natural language interface to an intrusion detection system. In *International Conference on Control, Automation and Systems, ICASE'01.*

Weir, David. 1988. *Characterizing mildly context-sensitive grammar formalisms.* Ph.D. thesis, University of Pennsylvania, Philadelphia.