Baby SRL: Modeling Early Language Acquisition.

Michael Connor
Department of Computer Science
University of Illinois
connor2@uiuc.edu

Yael Gertner
Beckman Institute
University of Illinois
ygertner@cyrus.psych.uiuc.edu

Cynthia Fisher
Department of Psychology
University of Illinois
cfisher@cyrus.psych.uiuc.edu

Dan Roth
Department of Computer Science
University of Illinois
danr@uiuc.edu

Abstract

A fundamental task in sentence comprehension is to assign semantic roles to sentence constituents. The structure-mapping account proposes that children start with a shallow structural analysis of sentences: children treat the number of nouns in the sentence as a cue to its semantic predicate-argument structure, and represent language experience in an abstract format that permits rapid generalization to new verbs. In this paper, we tested the consequences of these representational assumptions via experiments with a system for automatic semantic role labeling (SRL), trained on a sample of child-directed speech. When the SRL was presented with representations of sentence structure consisting simply of an ordered set of nouns, it mimicked experimental findings with toddlers, including a striking error found in children. Adding features representing the position of the verb increased accuracy and eliminated the error. We show the SRL system can use incremental knowledge gain to switch from error-prone noun order features to a more accurate representation, demonstrating a possible mechanism for this process in child development.

1 Introduction

How does the child get started in learning to interpret sentences? The structure-mapping view of early verb and syntax acquisition proposes that children start with a shallow structural analysis of sentences: children treat the number of nouns in the sentence as a cue to its semantic predicate-argument structure (Fisher, 1996), and represent language experience in an abstract format that permits rapid generalization to new verbs (Gertner et al., 2006).

The structure-mapping account makes strong predictions. First, as soon as children can identify some nouns, they should interpret transitive and intransitive sentences differently, simply by assigning a distinct semantic role to each noun in the sentence. Second, language-specific syntactic learning should transfer rapidly to new verbs. Third, some striking errors of interpretation can occur. In "Fred and Ginger danced", an intransitive verb is presented with two nouns. If children interpret any two-noun sentence as if it were transitive, they should be fooled into interpreting the order of two nouns in such conjoined-subject intransitive sentences as conveying agent-patient role information. Experiments with young children support these predictions. First, 21-month-olds use the number of nouns to understand sentences containing new verbs (Yuan et al., 2007). Second, 21-month-olds generalize what they have learned about English transitive word-order to sentences containing new verbs: Children who heard "The girl is gorping the boy" interpreted the girl as an agent and the boy as a patient (Gertner et al., 2006). Third, 21-month-olds make the predicted error, treating intransitive sentences containing two nouns as if they were transitive: they interpret the first noun in "The girl and the boy are gorping" as an agent and the second as a patient (Gertner and Fisher, 2006). This error is short-lived. By 25 months, children add new features to their representations of sentences, and interpret conjoined-subject intransitives differ-

CoNLL 2008: Proceedings of the 12th Conference on Computational Natural Language Learning, pages 81–88
Manchester, August 2008

ently from transitives (Naigles, 1990).

These experimental results shed light on what syntactic information children might have available for early sentence comprehension, but do not rule out the possibility that children's early performance is based on a more complex underlying system. In this paper, we tested the consequences of our representational assumptions by performing experiments with a system for automatic semantic role labeling (SRL), whose knowledge of sentence structure is under our control. Computational models of semantic role labeling learn to identify, for each verb in a sentence, all constituents that fill a semantic role, and to determine their roles. We adopt the architecture proposed by Roth and colleagues (Punyakanok et al., 2005), limiting the classifier's features to a set of lexical features and shallow structural features suggested by the structure-mapping account. Learning ability is measured by the level of SRL accuracy and, more importantly, the types of errors made by the system on sentences containing novel verbs. Testing these predictions on the automatic SRL provides us with a demonstration that it is possible to learn how to correctly assign semantic roles based only on these very simple cues.

From an NLP perspective this feature study provides evidence for the efficacy of alternative, simpler syntactic representations in gaining an initial foothold on sentence interpretation. It is clear that human learners do not begin interpreting sentences in possession of full part-of-speech tagging, or full parse trees. By building a model that uses shallow representations of sentences and mimics features of language development in children, we can explore the nature of initial representations of syntactic structure and build more complex features from there, further mimicking child development.

2 Learning Model

We trained a simplified SRL classifier (Baby SRL) with sets of features derived from the structure-mapping account. Our test used novel verbs to mimic sentences presented in experiments with children. Our learning task is similar to the full SRL task (Carreras and Màrquez, 2004), except that we classify the roles of individual words rather than full phrases. A full automatic SRL system (e.g. (Punyakanok et al., 2005)) typically involves multiple stages to 1) parse the input, 2) identify arguments, 3) classify those arguments, and then 4) run inference to make sure the final labeling for the full sentence does not violate any linguistic constraints. Our simplified SRL architecture (Baby SRL) essentially replaces the first two steps with heuristics. Rather than identifying arguments via a learned classifier with access to a full syntactic parse, the Baby SRL treats each noun in the sentence as a candidate argument and assigns a semantic role to it. A simple heuristic collapsed compound or sequential nouns to their final noun: an approximation of the head noun of the noun phrase. For example, 'Mr. Smith' was treated as the single noun 'Smith'. Other complex noun phrases were not simplified in this way. Thus, a phrase such as 'the toy on the floor' would be treated as two separate nouns, 'toy' and 'floor'. This represents the assumption that young children know 'Mr. Smith' is a single name, but they do not know all the predicating terms that may link multiple nouns into a single noun phrase. The simplified learning task of the Baby SRL implements a key assumption of the structure-mapping account: that at the start of multiword sentence comprehension children can tell which words in a sentence are nouns (Waxman and Booth, 2001), and treat each noun as a candidate argument.

Feedback is provided based on annotation in Propbank style: in training, each noun receives the role label of the phrase that noun is part of. Feedback is given at the level of the macro-role (agent, patient, etc., labeled A0-A4 for core arguments, and AM-* adjuncts). We also introduced a NO label for nouns that are not part of any argument.

For argument classification we use a linear classifier trained with a regularized perceptron update rule (Grove and Roth, 2001). This learning algorithm provides a simple and general linear classifier that has been demonstrated to work well in other text classification tasks, and allows us to inspect the weights of key features to determine their importance for classification. The Baby SRL does not use inference for the final classification. Instead it classifies every argument independently; thus multiple nouns can have the same role.

2.1 Training

The training data were samples of parental speech to one child ('Eve'; (Brown, 1973), available via Childes (MacWhinney, 2000)). We trained on parental utterances in samples 9 through 20, recorded at child age 21-27 months. All verb-containing utterances without symbols indicating long pauses or unintelligible words were automatically parsed with the Charniak parser (Charniak, 1997) and annotated using an existing SRL system (Punyakanok et al., 2005). In this initial pass, sentences with parsing errors that misidentified argument boundaries were excluded. Final role labels were hand-corrected using the Propbank annotation scheme (Kingsbury and Palmer, 2002). The child-directed speech (CDS) training set consisted of about 2200 sentences, of which a majority had a single verb and two nouns to be labeled[1]. We used the annotated CDS training data to train our Baby SRL, converting labeled phrases to labeled nouns in the manner described above.

3 Experimental Results

To evaluate the Baby SRL we tested it with sentences like those used for the experiments with children described above. All test sentences contained a novel verb ('gorp'). We constructed two test sentence templates: 'A gorps B' and 'A and B gorp', where A and B were replaced with nouns that appeared more than twice in training. We filled the A and B slots by sampling nouns that occurred roughly equally as the first and second of two nouns in the training data. This procedure was adopted to avoid 'building in' the predicted error by choosing A and B nouns biased toward an agent-patient interpretation. For each test sentence template we built a test set of 100 sentences by randomly sampling nouns in this fashion.

The test sentences with novel verbs ask whether the classifier transfers its learning about argument role assignment to unseen verbs. Does it assume the first of two nouns in a simple transitive sentence ('A gorps B') is the agent (A0) and the second is the patient (A1)? Does it overgeneralize this rule to two-noun intransitives ('A and B gorp'), mimicking children's behavior? We used two measures of success, one to assess classification accuracy, and the other to assess the predicted error. We used a per argument F1 for classification accuracy, with F1 based on correct identification of individual nouns rather than full phrases. Here precision is defined as the proportion of nouns that were given the correct label based on the argument they belong to, and recall is the proportion of complete arguments for which some noun in that argument was correctly labeled. The desired labeling for 'A gorps B' is A0 for the first argument and A1 for the second; for 'A and B gorp' both arguments should be A0. To measure predicted errors we also report the proportion of test sentences classified with A0 first and A1 second (%A0A1). This labeling is a correct generalization for the novel 'A gorps B' sentences, but is an overgeneralization for 'A and B gorp.'

3.1 Noun Pattern

The basic feature we propose is the noun pattern feature. We hypothesize that children use the number and order of nouns to represent argument structure. To encode this we created a feature (NPattern) that indicates how many nouns there are in the sentence and which noun the target is. For example, in our two-noun test sentences noun A has the feature '_N' active indicating that it is the first noun of two. Likewise for B the feature 'N_' is active, indicating that it is the second of two nouns. This feature is easy to compute once nouns are identified, and does not require fine-grained distinctions between types of nouns or any other part of speech. Table 1 shows the initial feature progression that involves this feature. The baseline system (feature set 1) uses lexical features only: the target noun and the root form of the predicate.

We first tested the hypothesis that children use the NPattern features to distinguish different noun arguments, but only for specific verbs. The NPattern&V features are conjunctions of the target verb and the noun pattern, and these are added to the word features to form feature set 2. Now every example has three features active: target noun, target predicate, and a NPattern&V feature indicating 'the target is the first of two nouns and the verb is X.' This feature does not improve results on the novel 'A gorps B' test set, or generate the predicted error with the 'A and B gorp' test set, because the verb-specific NPattern&V features provide no way to generalize to unseen verbs.

We next tested the NPattern feature alone, without making it verb-specific (feature set 3). The noun pattern feature was added to the word features and again each example had three features active: target noun, target predicate, and the target's noun-pattern feature (first of two, second of three, etc.). The abstract NPattern feature allows the Baby SRL to generalize to new verbs: it increases the system's tendency to predict that the first of two

[1]Corpus available at `http://L2R.cs.uiuc.edu/~cogcomp/data.php`

	CHILDES								WSJ			
	Unbiased Noun Choice				Biased Noun Choice				Biased Noun Choice			
	A gorps B		A and B gorp		A gorps B		A and B gorp		A gorps B		A and B gorp	
Features	F1	%A0A1	F1	%A0A1	F1	%A0A1	F1	%A0A1	F1	%A0A1	F1	%A0A1
1. Words	0.59	0.38	0.46	0.38	0.80	0.65	0.53	0.65	0.57	0.31	0.37	0.31
2. NPattern&V	0.53	0.28	0.54	0.28	0.81	0.67	0.53	0.67	0.56	0.31	0.39	0.31
3. NPattern	0.83	0.65	0.33	0.65	0.96	0.92	0.46	0.92	0.67	0.44	0.37	0.44
4. NPattern + NPattern&V	0.83	0.65	0.33	0.65	0.95	0.90	0.45	0.90	0.73	0.53	0.44	0.53
5. + VPosition	0.99	0.96	0.98	0.00	1.00	1.00	0.99	0.01	0.94	0.88	0.69	0.39

Table 1: *Experiments showing the efficacy of Noun Pattern features for determining agent/patient roles in simple two-noun sentences. The novel verb test sets assess whether the Baby SRL generalizes transitive argument prediction to unseen verbs in the case of 'A gorps B' (increasing %A0A1 and thus F1), and overgeneralizes in the case of 'A and B gorp' (increasing %A0A1, which is an error). By varying the sampling method for creating the test sentences we can start with a biased or unbiased lexical baseline, demonstrating that the noun pattern features still improve over knowledge that can be contained in typical noun usage. The simple noun pattern features are still effective at learning this pattern when trained with more complex Wall Street Journal training data.*

nouns is A0 and the second of two nouns is A1 for verbs not seen in training. Feature set 4 includes both the abstract, non-verb-specific NPattern feature and the verb-specific version. This feature set preserves the ability to generalize to unseen verbs; thus the availability of the verb-specific NPattern features during training did not prevent the abstract NPattern features from gathering useful information.

3.2 Lexical Cues for Role-Labeling

Thus far, the target nouns' lexical features provided little help in role labeling, allowing us to clearly see the contribution of the proposed simple structural features. Would our structural features produce any improvement above a more realistic lexical baseline? We created a new set of test sentences, sampling the A nouns based on the distribution of nouns seen as the first of two nouns in training, and the B nouns based on the distribution of nouns seen as the second of two nouns. Given this revised sampling of nouns, the words-only baseline is strongly biased toward A0A1 (biased results for feature set 1 in table 1). This high baseline reflects a general property of conversation: Lexical choices provide considerable information about semantic roles. For example, the 6 most common nouns in the Eve corpus are pronouns that are strongly biased in their positions and in their semantic roles (e.g., 'you', 'it'). Despite this high baseline, however, we see the same pattern in the unbiased and biased experiments in table 1. The addition of the NPattern features (feature set 3) substantially improves performance on 'A gorps B' test sentences, and promotes overgeneralization errors on 'A and B gorp' sentences.

3.3 More Complex Training Data

For comparison purposes we also trained the Baby SRL on a subset of the Propbank training data of Wall Street Journal (WSJ) text (Kingsbury and Palmer, 2002). To approximate the simpler sentences of child-directed speech we selected only those sentences with 8 or fewer words. This provided a training set of about 2500 sentences, most with a single verb and two nouns to be labeled. The CDS and WSJ data pose similar problems for learning abstract and verb-specific knowledge. However, newspaper text differs from casual speech to children in many ways, including vocabulary and sentence complexity. One could argue that the WSJ corpus presents a worst-case scenario for learning based on shallow representations of sentence structure: Full passive sentences are more common in written corpora such as the WSJ than in samples of conversational speech, for example (Roland et al., 2007). As a result of such differences, two-noun sequences are less likely to display an A0-A1 sequence in the WSJ (0.42 A0-A1 in 2-noun sentences) than in the CDS training data (0.67 A0-A1). The WSJ data provides a more demanding test of the Baby SRL.

We trained the Baby SRL on the WSJ data, and tested it using the biased lexical choices as described above, sampling A and B nouns for novel-verb test sentences based on the distribution of nouns seen as the first of two nouns in training, and as the second of two nouns, respectively. The WSJ training produced performance strikingly similar to the performance resulting from CDS training (last 4 columns of Table 1). Even in this more complex training set, the addition of the NPattern features (feature set 3) improves performance on 'A gorps B' test sentences, and promotes over-generalization errors on 'A and B gorp' sentences.

3.4 Tests with Familiar Verbs

Features	Total	A0	A1	A2	A4
1. Words	0.64	0.83	0.74	0.33	0.00
2. NPattern&V	0.67	0.86	0.77	0.45	0.44
3. NPattern	0.66	0.87	0.76	0.37	0.22
4. NPattern + NPattern&V	0.68	0.87	0.80	0.47	0.44
5. + VPosition	0.70	0.88	0.83	0.50	0.50

Table 2: *Testing NPattern features on full SRL task of heldout section 8 of Eve when trained on sections 9 through 20. Each result column reflects a per argument F1.*

Learning to interpret sentences depends on balancing abstract and verb-specific structural knowledge. Natural linguistic corpora, including our CDS training data, have few verbs of very high frequency and a long tail of rare verbs. Frequent verbs occur with differing argument patterns. For example, 'have' and 'put' are frequent in the CDS data. 'Have' nearly always occurs in simple transitive sentences that display the canonical word order of English (e.g., 'I have cookies'). 'Put', in contrast, tends to appear in non-canonical sentences that do not display an agent-patient ordering, including imperatives ('Put it on the floor'). To probe the Baby SRL's ability to learn the argument-structure preferences of familiar verbs, we tested it on a held-out sample of CDS from the same source (Eve sample 8, approximately 234 labeled sentences). Table 2 shows the same feature progression shown previously, with the full SRL test set. The words-only baseline (feature set 1 in Table 2) yields fairly accurate performance, showing that considerable success in role assignment in these simple sentences can be achieved based on the argument-role biases of the target nouns and the familiar verbs. Despite this high baseline, however, we still see the benefit of simple structural features. Adding verb-specific (feature set 2) or abstract NPattern features (feature set 3) improves classification performance, and the combination of both verb-specific and abstract NPattern features (feature set 4) yields higher performance than either alone. The combination of abstract NPattern features with the verb-specific versions allows the Baby SRL both to generalize to unseen verbs, as seen in earlier sections, and to learn the idiosyncrasies of known verbs.

3.5 Verb Position

The noun pattern feature results show that the Baby SRL can learn helpful rules for argument-role assignment using only information about the number and order of nouns. It also makes the error predicted by the structure-mapping account, and documented in children, because it has no way to represent the difference between the 'A gorps B' and 'A and B gorp' test sentences. At some point the learner must develop more sophisticated syntactic representations that could differentiate these two. These could include many aspects of the sentence, including noun-phrase and verb-phrase morphological features, and word-order features. As a first step in examining recovery from the predicted error, we focused on word-order features. We did this by adding a verb position feature (VPosition) that specifies whether the target noun is before or after the verb. Now simple transitive sentences in training should support the generalization that preverbal nouns tend to be agents, and post-verbal nouns tend to be patients. In testing, the Baby SRL's classification of the 'A gorps B' and 'A and B gorp' sentences should diverge.

When we add verb position information (feature set 5 in table 1 and 2), performance improves still further for transitive sentences, both with biased and unbiased test sentences. Also, for the first time, the A0A1 pattern is predicted less often for 'A and B gorp' sentences. This error diminished because the classifier was able to use the verb position features to distinguish these from 'A gorps B' sentences.

	Unbiased Lexical			
	A gorps B		A and B gorp	
Features	F1	%A0A1	F1	%A0A1
1. Words	0.59	0.38	0.46	0.38
3. NPattern	0.83	0.65	0.33	0.65
6. VPosition	0.99	0.95	0.97	0.00

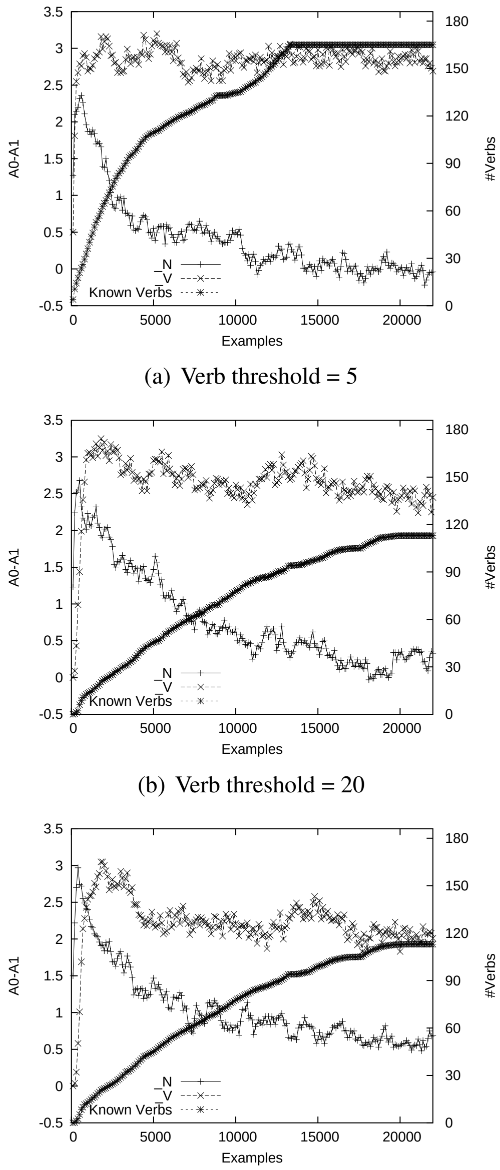
Table 3: *Verb Position vs. Noun Pattern features alone. Verb position features yield better overall performance, but do not replicate the error on 'A and B gorp' sentences seen with children.*

Verb position alone provides another simple abstract representation of sentence structure, so it might be proposed as an equally natural initial representation for human learners, rather than the noun pattern features we proposed. The VPosition features should also support learning and generalization of word-order rules for interpreting transitive sentences, thus reproducing some of the data from children that we reviewed above. In table 3 we compared the words-only baseline (set 1), words and NPattern features (set 3), and a new feature set, words and VPosition (set 6). In terms of correct performance on novel transitive verbs ('A gorps B'), the VPosition features outperform the NPattern features. This may be partly because the same VPosition features are used in all sentences during training, while the NPattern features partition sentences by number of nouns, but is also due to the fact that the verb position features provide a more sophisticated representation of English sentence structure. Verb position features can distinguish transitive sentences from imperatives containing multiple post-verbal nouns, for example. Although verb position is ultimately a more powerful representation of word order for English sentences, it does not accurately reproduce a 21-month-old's performance on all aspects of this task. In particular, the VPosition feature does not support the overgeneralization of the A0A1 pattern to the 'A and B gorp' test sentences. This suggests that children's very early sentence comprehension is dominated by less sophisticated representations of word order, akin to the NPattern features we proposed.

3.6 Informativeness vs. Availability

In the preceding sections, we modeled increases in syntactic knowledge by building in more sophisticated features. The Baby SRL escaped the predicted error on two-noun intransitive sentences when given access to features reflecting the position of the target noun relative to the verb. This imposed sequence of features is useful as a starting point, but a more satisfying approach would be to use the Baby SRL to explore possible reasons why NPattern features might dominate early in acquisition, even though VPosition features are ultimately more useful for English.

In theory, a feature might be unavailable early in acquisition because of its computational complexity. For example, lexical features are presumably less complex than relative position features such as NPattern and VPosition. In practice, features can also be unavailable at first because of an informational lack. Here we suggest that NPattern features might dominate VPosition features early in acquisition because the early lexicon is dominated by nouns, and it is easier to compute position relative to a known word than to an unknown word. Many studies have shown that children's early vocabulary is dominated by names for objects and people (Gentner and Boroditsky, 2001).



(a) Verb threshold = 5

(b) Verb threshold = 20

(c) Verb threshold = 20, +verb-specific features

Figure 1: *Testing the consequences of the assumption that Verb Position features are only active for familiar verbs. The figure plots the bias of the features '_N' and '_V' to predict A0 over A1, as the difference between the weights of these connections in the learned network. Verb position features win out over noun pattern features as the verb vocabulary grows. Varying the verb familiarity threshold ((a) vs. (b)) and the presence versus absence of verb-specific versions of the structural features ((b) vs. (c)) affects how quickly the verb position features become dominant.*

To test the consequences of this proposed informational bottleneck on the relative weighting of NPattern and VPosition features during training, we modified the Baby SRL's training procedure such that NPattern features were always active, but VPosition features were active during training only when the verb in the current example had been encountered a critical number of times. This represents the assumption that the child can recognize which words in the sentence are nouns, based on lexical familiarity or morphological context (Waxman and Booth, 2001), but is less likely to be able to represent position relative to the verb without knowing the verb well.

Figure 1 shows the tendency of the NPattern feature '_N' (first of two nouns) and the VPosition feature '_V' (pre-verbal noun) to predict the role A0 as opposed to A1 as the difference between the weights of these connections in the learned network. Figure 1(a) shows the results when VPosition features were active whenever the target verb had occurred at least 5 times; in Figure 1(b) the threshold for verb familiarity was 20. In both figures we see that the VPosition features win out over the NPattern features as the verb vocabulary grows. Varying the degree of verb familiarity required to accurately represent VPosition features affects how quickly the VPosition features win out (compare Figures 1(a) and 1(b)). Figure 1(c) shows the same analysis with a threshold of 20, but with verb-specific as well as abstract versions of the NPattern and the VPosition features. In this procedure, every example started with three features: target noun, target predicate, NPattern, and if the verb was known, added NPattern&V, VPosition, and VPosition&V. Comparing Figures 1(b) and 1(c), we see that the addition of verb-specific versions of the structural features also affects the rate at which the VPosition features come to dominate the NPattern features.

Thus, in training the VPosition features become dominant as the SRL learns to recognize more verbs. However, the VPosition features are inactive when the Baby SRL encounters the novel-verb test sentences. Since the NPattern features are active in test, the system generates the predicted error until the bias of the NPattern features reaches 0. Note in figure 1(c) that when verb-specific structural features were added, the Baby SRL never learned to entirely discount the NPattern features within the range of training provided. This result is reminiscent of suggestions in the psycholinguistics literature that shallow representations of syntax persist in the adult parser, alongside more sophisticated representations (e.g., (Ferreira, 2003)).

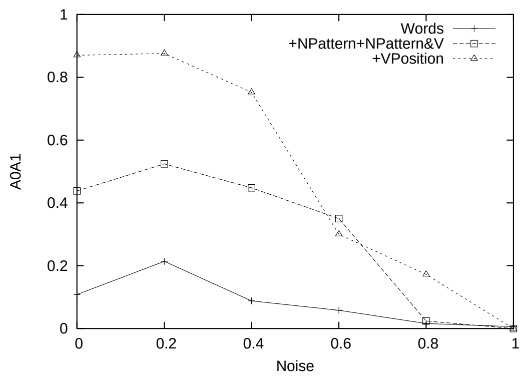


Figure 2: *Testing the ability of simple features to cope with varying amounts of noisy feedback. Even with noisy feedback, the noun pattern features support learning and generalization to new verbs of a simple agent-patient template for understanding transitive sentences. These results are lower than those found in table 1 due to slightly different training assumptions.*

3.7 Noisy Training

So far, the Baby SRL has only been trained with perfect feedback. Theories of human language acquisition assume that learning to understand sentences is naturally a partially-supervised task: the child uses existing knowledge of words and syntax to assign a meaning to a sentence; the appropriateness of this meaning for the referential context provides the feedback (e.g., (Pinker, 1989)). But this feedback must be noisy. Referential scenes provide useful but often ambiguous information about the semantic roles of sentence participants. For example, a participant could be construed as an agent of fleeing or as a patient being chased. In a final set of experiments, we examined the generalization abilities of the Baby SRL as a function of the integrity of semantic feedback.

We provided noisy semantic-role feedback during training by giving a randomly-selected argument label on 0 to 100% of examples. Following this training, we tested with the 'A gorps B' test sentences, using the unbiased noun choices.

As shown in Figure 2, feature sets including NPattern or VPosition features yield reasonable performance on the novel verb test sentences up to 50% noise, and promote an A0-A1 sequence over the words-only baseline even at higher noise levels. Thus the proposed simple structural features are robust to noisy feedback.

4 Conclusion

The simplified SRL classifier mimicked experimental results with toddlers. We structured the learning task to ask whether shallow representations of sentence structure provided a useful initial representation for learning to interpret sentences. Given representations of the number and order of nouns in the sentence (noun pattern features), the Baby SRL learned to classify the first of two nouns as an agent and the second as a patient. When provided with both verb-general and verb-specific noun pattern features, the Baby SRL learned to balance verb-specific and abstract syntactic knowledge. By treating each noun as an argument, it also reproduced the errors children make. Crucially, verb-position features improved performance when added to the noun-pattern feature, but when presented alone failed to produce the error found with toddlers. We believe that our model can be naturally extended to support the case in which the arguments are noun phrases rather than single noun words and this extension is one of the first steps we will explore next.

Acknowledgments

We would like to thank our annotators, especially Yuancheng Tu. This research is supported by NSF grant BCS-0620257 and NIH grant R01-HD054448.

References

Brown, R. 1973. *A First Language*. Harvard University Press, Cambridge, MA.

Carreras, X. and L. Màrquez. 2004. Introduction to the CoNLL-2004 shared tasks: Semantic role labeling. In *Proceedings of CoNLL-2004*, pages 89–97. Boston, MA, USA.

Charniak, E. 1997. Statistical parsing with a context-free grammar and word statistics. In *Proc. National Conference on Artificial Intelligence*.

Ferreira, F. 2003. The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47:164–203.

Fisher, C. 1996. Structural limits on verb mapping: The role of analogy in children's interpretation of sentences. *Cognitive Psychology*, 31:41–81.

Gentner, D. and L. Boroditsky. 2001. Individuation, relativity and early word learning. In Bowerman, M. and S. C. Levinson, editors, *Language acquisition and conceptual development*, pages 215–256. Cambridge University Press, New York.

Gertner, Y. and C. Fisher. 2006. Predicted errors in early verb learning. In *31st Annual Boston University Conference on Language Development*.

Gertner, Y., C. Fisher, and J. Eisengart. 2006. Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, 17:684–691.

Grove, A. and D. Roth. 2001. Linear concepts and hidden variables. *Machine Learning*, 42(1/2):123–141.

Kingsbury, P. and M. Palmer. 2002. From Treebank to PropBank. In *Proceedings of LREC-2002*, Spain.

MacWhinney, B. 2000. *The CHILDES project: Tools for analyzing talk. Third Edition*. Lawrence Elrbaum Associates, Mahwah, NJ.

Naigles, L. R. 1990. Children use syntax to learn verb meanings. *Journal of Child Language*, 17:357–374.

Pinker, S. 1989. *Learnability and Cognition*. Cambridge: MIT Press.

Punyakanok, V., D. Roth, and W. Yih. 2005. The necessity of syntactic parsing for semantic role labeling. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1117–1123.

Roland, D., F. Dick, and J. L. Elman. 2007. Frequency of basic english grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57:348–379.

Waxman, S. R. and A. Booth. 2001. Seeing pink elephants: Fourteen-month-olds's interpretations of novel nouns and adjectives. *Cognitive Psychology*, 43:217–242.

Yuan, S., C. Fisher, Y. Gertner, and J. Snedeker. 2007. Participants are more than physical bodies: 21-month-olds assign relational meaning to novel transitive verbs. In *Biennial Meeting of the Society for Research in Child Development*, Boston, MA.